

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 24, 2004, 15:33:13 ; Search time 16.4552 Seconds  
(without alignments)  
47.060 Million cell updates/sec

Title: US-09-641-801-8  
Perfect score: 82  
Sequence: 1 LKPFPKLKVEVFPFP 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

		8					
Result		Query					
No.	Score	Match	Length	DB	ID		Description
1	82	100.0	15	4	US-09-641-803-8		Sequence 8, Appli
2	40	48.8	451	4	US-09-555-820A-16		Sequence 16, Appl
3	40	48.8	462	4	US-09-555-820A-20		Sequence 20, Appl
4	40	48.8	618	4	US-09-555-820A-14		Sequence 14, Appl
5	40	48.8	626	4	US-09-214-619-4		Sequence 4, Appli
6	40	48.8	626	4	US-09-555-820A-12		Sequence 12, Appl
7	39	47.6	346	4	US-09-328-352-5541		Sequence 5541, Ap
8	39	47.6	642	4	US-09-328-352-5199		Sequence 5199, Ap
9	39	47.6	759	2	US-08-637-759B-89		Sequence 89, Appl
10	39	47.6	759	3	US-08-871-355A-89		Sequence 89, Appl
11	39	47.6	759	4	US-09-201-945-89		Sequence 89, Appl

12	38	46.3	164	4	US-09-621-976-5565	Sequence 5565, Ap
13	38	46.3	178	4	US-09-328-352-7054	Sequence 7054, Ap
14	38	46.3	312	3	US-08-793-331-2	Sequence 2, Appli
15	38	46.3	724	3	US-08-793-331-4	Sequence 4, Appli
16	38	46.3	725	3	US-08-793-331-6	Sequence 6, Appli
17	38	46.3	1536	3	US-09-413-814-10	Sequence 10, Appl
18	37	45.1	71	4	US-09-328-352-5884	Sequence 5884, Ap
19	37	45.1	162	5	PCT-US93-03076-8	Sequence 8, Appli
20	37	45.1	191	4	US-09-134-000C-6524	Sequence 6524, Ap
21	37	45.1	231	4	US-09-107-532A-5301	Sequence 5301, Ap
22	37	45.1	486	4	US-09-252-991A-30297	Sequence 30297, A
23	37	45.1	517	4	US-09-252-991A-21294	Sequence 21294, A
24	37	45.1	943	2	US-08-808-982-7	Sequence 7, Appli
25	37	45.1	943	3	US-09-306-902A-7	Sequence 7, Appli
26	37	45.1	1253	4	US-08-864-785-2	Sequence 2, Appli
27	37	45.1	1513	5	PCT-US93-03076-2	Sequence 2, Appli
28	37	45.1	2756	1	US-08-375-709-11	Sequence 11, Appl
29	37	45.1	2756	1	US-08-752-929-11	Sequence 11, Appl
30	37	45.1	2756	3	US-09-090-793-7	Sequence 7, Appli
31	37	45.1	2756	4	US-09-231-899-7	Sequence 7, Appli
32	36	43.9	18	4	US-09-641-803-22	Sequence 22, Appl
33	36	43.9	114	4	US-09-198-452A-116	Sequence 116, App
34	36	43.9	209	3	US-09-269-220-1	Sequence 1, Appli
35	36	43.9	209	3	US-09-269-220-2	Sequence 2, Appli
36	36	43.9	209	4	US-08-836-778-2	Sequence 2, Appli
37	36	43.9	222	2	US-08-391-743A-2	Sequence 2, Appli
38	36	43.9	222	4	US-09-143-155-2	Sequence 2, Appli
39	36	43.9	312	3	US-09-230-637-34	Sequence 34, Appl
40	36	43.9	313	4	US-09-347-878-30	Sequence 30, Appl
41	36	43.9	313	4	US-09-367-007C-39	Sequence 39, Appl
42	36	43.9	343	1	US-08-180-209B-56	Sequence 56, Appl
43	36	43.9	343	3	US-08-474-853-56	Sequence 56, Appl
44	36	43.9	343	4	US-09-166-205B-56	Sequence 56, Appl
45	36	43.9	343	5	PCT-US94-02629-56	Sequence 56, Appl

# ALIGNMENTS

## RESULT 1

US-09-641-803-8

; Sequence 8, Application US/09641803

; Patent No. 6500798

; GENERAL INFORMATION:

; APPLICANT: STANTON, G. John

; APPLICANT: HUGHES, Thomas K.

; APPLICANT: BOLDOGH, Istvan

; TITLE OF INVENTION: USE OF COLOSTRININ, CONSTITUENT PEPTIDES THEREOF, AND

; TITLE OF INVENTION: ANALOGS THEREOF AS OXIDATIVE STRESS REGULATORS

; FILE REFERENCE: 265.00220101

; CURRENT APPLICATION NUMBER: US/09/641,803

; CURRENT FILING DATE: 2000-08-17

; PRIOR APPLICATION NUMBER: 60/149,310

; PRIOR FILING DATE: 1999-08-17

; NUMBER OF SEQ ID NOS: 34

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 8

; LENGTH: 15  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: peptide  
US-09-641-803-8

Query Match 100.0%; Score 82; DB 4; Length 15;  
Best Local Similarity 100.0%; Pred. No. 4.7e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LKPFPKLKVEVFPPF 15  
| | | | | | | | | | | | | | |  
Db 1 LKPFPKLKVEVFPPF 15

#### RESULT 2

US-09-555-820A-16  
; Sequence 16, Application US/09555820A  
; Patent No. 6680429  
; GENERAL INFORMATION:  
; APPLICANT: Webster, David  
; APPLICANT: Burgess, Diane  
; TITLE OF INVENTION: A Starchless Variety of Pisum Sativum Having Elevated Levels of Sucrose  
; FILE REFERENCE: SVS3801P0302US  
; CURRENT APPLICATION NUMBER: US/09/555,820A  
; CURRENT FILING DATE: 2000-08-29  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 16  
; LENGTH: 451  
; TYPE: PRT  
; ORGANISM: Pisum sativum  
US-09-555-820A-16

Query Match 48.8%; Score 40; DB 4; Length 451;  
Best Local Similarity 63.6%; Pred. No. 83;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 4 FPKLKVEVFPPF 14  
| | | | : | | |  
Db 31 FPSFKVQNFPF 41

#### RESULT 3

US-09-555-820A-20  
; Sequence 20, Application US/09555820A  
; Patent No. 6680429  
; GENERAL INFORMATION:  
; APPLICANT: Webster, David  
; APPLICANT: Burgess, Diane  
; TITLE OF INVENTION: A Starchless Variety of Pisum Sativum Having Elevated Levels of Sucrose  
; FILE REFERENCE: SVS3801P0302US  
; CURRENT APPLICATION NUMBER: US/09/555,820A

; CURRENT FILING DATE: 2000-08-29  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 20  
; LENGTH: 462  
; TYPE: PRT  
; ORGANISM: Pisum sativum  
US-09-555-820A-20

Query Match 48.8%; Score 40; DB 4; Length 462;  
Best Local Similarity 63.6%; Pred. No. 85;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 4 FPKLKVEVFPP 14  
|| ||: |||  
Db 33 FPSFKVQNFPF 43

RESULT 4

US-09-555-820A-14  
; Sequence 14, Application US/09555820A  
; Patent No. 6680429  
; GENERAL INFORMATION:  
; APPLICANT: Webster, David  
; APPLICANT: Burgess, Diane  
; TITLE OF INVENTION: A Starchless Variety of Pisum Sativum Having Elevated Levels of Sucrose  
; FILE REFERENCE: SVS3801P0302US  
; CURRENT APPLICATION NUMBER: US/09/555,820A  
; CURRENT FILING DATE: 2000-08-29  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 14  
; LENGTH: 618  
; TYPE: PRT  
; ORGANISM: Pisum sativum  
US-09-555-820A-14

Query Match 48.8%; Score 40; DB 4; Length 618;  
Best Local Similarity 63.6%; Pred. No. 1.1e+02;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 4 FPKLKVEVFPP 14  
|| ||: |||  
Db 33 FPSFKVQNFPF 43

RESULT 5

US-09-214-619-4  
; Sequence 4, Application US/09214619  
; Patent No. 6538180  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: METHOD FOR INCREASING SUCROSE  
; TITLE OF INVENTION: CONTENT OF PLANTS  
; NUMBER OF SEQUENCES: 4  
; COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/214,619
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 626 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-214-619-4

```

```

Query Match          48.8%; Score 40; DB 4; Length 626;
Best Local Similarity 63.6%; Pred. No. 1.2e+02;
Matches      7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy      4 FPKLKVEVFPPF 14
        || ||: |||
Db      33 FPSFKVQNFPF 43

```

```

RESULT 6
US-09-555-820A-12
; Sequence 12, Application US/09555820A
; Patent No. 6680429
; GENERAL INFORMATION:
; APPLICANT: Webster, David
; APPLICANT: Burgess, Diane
; TITLE OF INVENTION: A Starchless Variety of Pisum Sativum Having Elevated
Levels of Sucrose
; FILE REFERENCE: SVS3801P0302US
; CURRENT APPLICATION NUMBER: US/09/555,820A
; CURRENT FILING DATE: 2000-08-29
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 626
; TYPE: PRT
; ORGANISM: Pisum sativum
US-09-555-820A-12

```

```

Query Match          48.8%; Score 40; DB 4; Length 626;
Best Local Similarity 63.6%; Pred. No. 1.2e+02;
Matches      7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy      4 FPKLKVEVFPPF 14
        || ||: |||
Db      33 FPSFKVQNFPF 43

```

```

RESULT 7
US-09-328-352-5541
; Sequence 5541, Application US/09328352
; Patent No. 6562958

```

```

; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5541
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5541

```

```

Query Match          47.6%; Score 39; DB 4; Length 346;
Best Local Similarity 88.9%; Pred. No. 92;
Matches      8; Conservative    0; Mismatches    1; Indels      0; Gaps      0;

```

```

Qy      1 LKPFPKLV 9
        || |||||
Db      164 LKQFPKLV 172

```

```

RESULT 8
US-09-328-352-5199
; Sequence 5199, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5199
; LENGTH: 642
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5199

```

```

Query Match          47.6%; Score 39; DB 4; Length 642;
Best Local Similarity 46.7%; Pred. No. 1.7e+02;
Matches      7; Conservative    2; Mismatches    6; Indels      0; Gaps      0;

```

```

Qy      1 LKPFPKLVKVEVFPFP 15
        ||| | :| :| |
Db      140 LKPLSKQLIEQYPLP 154

```

```

RESULT 9
US-08-637-759B-89
; Sequence 89, Application US/08637759B
; Patent No. 5876931
; GENERAL INFORMATION:

```

```

;   APPLICANT:  David William Holden
;   TITLE OF INVENTION:  Identification of Genes
;   NUMBER OF SEQUENCES:  501
;   CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Patrea L. Pabst
;       STREET:  2800 One Atlantic Center
;       STREET:  1201 West Peachtree Street
;       CITY:  Atlanta
;       STATE:  Georgia
;       COUNTRY:  USA
;       ZIP:  30309-3450
;   COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/637,759B
;       FILING DATE:  03-MAY-1996
;       CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/GB95/02875
;       FILING DATE:  11-DEC-1995
;       CLASSIFICATION:  435
;   ATTORNEY/AGENT INFORMATION:
;       NAME:  Pabst, Patrea L.
;       REGISTRATION NUMBER:  31,284
;       REFERENCE/DOCKET NUMBER:  RPMS 101
;   TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (404) 873-8794
;       TELEFAX:  (404) 873-8795
;   INFORMATION FOR SEQ ID NO:  89:
;       SEQUENCE CHARACTERISTICS:
;           LENGTH:  759 amino acids
;           TYPE:  amino acid
;           STRANDEDNESS:  single
;           TOPOLOGY:  linear
;       MOLECULE TYPE:  protein
;       HYPOTHETICAL:  NO
US-08-637-759B-89

```

```

Query Match          47.6%;  Score 39;  DB 2;  Length 759;
Best Local Similarity 46.2%;  Pred. No. 2.1e+02;
Matches      6;  Conservative      3;  Mismatches      4;  Indels      0;  Gaps      0;

```

```

Qy          3 PFPKLVKVEVFPFP 15
             | |:::|| | |
Db          462 PLPEVNIEVLPEP 474

```

# RESULT 10

US-08-871-355A-89

```

; Sequence 89, Application US/08871355A
; Patent No. 6015669
; GENERAL INFORMATION:
;   APPLICANT:  David William Holden
;   TITLE OF INVENTION:  Identification of Genes

```

```

;   NUMBER OF SEQUENCES:  501
;   CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Patrea L. Pabst
;       STREET:  2800 One Atlantic Center
;       STREET:  1201 West Peachtree Street
;       CITY:  Atlanta
;       STATE:  Georgia
;       COUNTRY:  USA
;       ZIP:  30309-3450
;   COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/871,355A
;       FILING DATE:  09-JUN-1997
;       CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/GB95/02875
;       FILING DATE:  11-DEC-1995
;       CLASSIFICATION:  435
;   ATTORNEY/AGENT INFORMATION:
;       NAME:  Pabst, Patrea L.
;       REGISTRATION NUMBER:  31,284
;       REFERENCE/DOCKET NUMBER:  RPMS 101 CON
;   TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (404) 873-8794
;       TELEFAX:  (404) 873-8795
;   INFORMATION FOR SEQ ID NO:  89:
;       SEQUENCE CHARACTERISTICS:
;           LENGTH:  759 amino acids
;           TYPE:  amino acid
;           STRANDEDNESS:  single
;           TOPOLOGY:  linear
;       MOLECULE TYPE:  protein
;       HYPOTHETICAL:  NO
US-08-871-355A-89

```

```

Query Match          47.6%;  Score 39;  DB 3;  Length 759;
Best Local Similarity 46.2%;  Pred. No. 2.1e+02;
Matches      6;  Conservative      3;  Mismatches      4;  Indels      0;  Gaps      0;

```

```

Qy      3 PFPKLVKVEVFPFP 15
        | |:::|| | |
Db      462 PLPEVNIEVLPEP 474

```

```

RESULT 11
US-09-201-945-89
; Sequence 89, Application US/09201945
; Patent No. 6342215
;   GENERAL INFORMATION:
;       APPLICANT:  David William Holden
;       TITLE OF INVENTION:  Identification of Genes
;       NUMBER OF SEQUENCES:  501
;       CORRESPONDENCE ADDRESS:

```



; ADDRESSEE: Patrea L. Pabst  
 ; STREET: 2800 One Atlantic Center  
 ; STREET: 1201 West Peachtree Street  
 ; CITY: Atlanta  
 ; STATE: Georgia  
 ; COUNTRY: USA  
 ; ZIP: 30309-3450  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/201,945  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/637,759  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Pabst, Patrea L.  
 ; REGISTRATION NUMBER: 31,284  
 ; REFERENCE/DOCKET NUMBER: RPMS 101  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (404) 873-8794  
 ; TELEFAX: (404) 873-8795  
 ; INFORMATION FOR SEQ ID NO: 89:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 759 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; HYPOTHETICAL: NO  
 US-09-201-945-89

Query Match 47.6%; Score 39; DB 4; Length 759;  
 Best Local Similarity 46.2%; Pred. No. 2.1e+02;  
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 3 PFPKLVKVEVFPFP 15  
 | |:::|| | |  
 Db 462 PLPEVNIEVLPEP 474

# RESULT 12

US-09-621-976-5565  
 ; Sequence 5565, Application US/09621976  
 ; Patent No. 6639063  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dumas Milne Edwards, J.B.  
 ; APPLICANT: Jobert, S.  
 ; APPLICANT: Giordano, J.Y.  
 ; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
 ; FILE REFERENCE: GENSET.054PR2  
 ; CURRENT APPLICATION NUMBER: US/09/621,976

; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 5565  
; LENGTH: 164  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: -16..-1  
US-09-621-976-5565

Query Match 46.3%; Score 38; DB 4; Length 164;  
Best Local Similarity 87.5%; Pred. No. 61;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 8 KVEVFPPF 15  
| |||||  
Db 50 KKEVFPPF 57

RESULT 13

US-09-328-352-7054

; Sequence 7054, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 7054  
; LENGTH: 178  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-7054

Query Match 46.3%; Score 38; DB 4; Length 178;  
Best Local Similarity 53.8%; Pred. No. 67;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 LKPFPKLKVEVFP 13  
||||| : :  
Db 28 LKPFPIAIDNIP 40

RESULT 14

US-08-793-331-2

; Sequence 2, Application US/08793331  
; Patent No. 6071877  
; GENERAL INFORMATION:  
; APPLICANT: DELECLUSE, ARMELLE  
; APPLICANT: THIERY, ISABELLE  
; TITLE OF INVENTION: NEW POLYPEPTIDES HAVING A TOXIC ACTIVITY AGAINST

```
; TITLE OF INVENTION: INSECTS OF THE DIPTERAE FAMILY
; FILE REFERENCE: 0660-0116-0 PCT
; CURRENT APPLICATION NUMBER: US/08/793,331
; CURRENT FILING DATE: 1997-05-13
; EARLIER APPLICATION NUMBER: PCT/FR95/01116
; EARLIER FILING DATE: 1995-08-24
; EARLIER APPLICATION NUMBER: FR 94/10299
; EARLIER FILING DATE: 1994-08-25
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 312
; TYPE: PRT
; ORGANISM: B. thuringiensis ser. jegathesan
US-08-793-331-2
```

```
Query Match          46.3%; Score 38; DB 3; Length 312;
Best Local Similarity 70.0%; Pred. No. 1.2e+02;
Matches      7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy      4 FPKLKVEVFP 13
        | ||| |:|
Db      55 FAKLKSEIFP 64
```

```
RESULT 15
US-08-793-331-4
; Sequence 4, Application US/08793331
; Patent No. 6071877
; GENERAL INFORMATION:
; APPLICANT: DELECLUSE, ARMELLE
; APPLICANT: THIERY, ISABELLE
; TITLE OF INVENTION: NEW POLYPEPTIDES HAVING A TOXIC ACTIVITY AGAINST
; TITLE OF INVENTION: INSECTS OF THE DIPTERAE FAMILY
; FILE REFERENCE: 0660-0116-0 PCT
; CURRENT APPLICATION NUMBER: US/08/793,331
; CURRENT FILING DATE: 1997-05-13
; EARLIER APPLICATION NUMBER: PCT/FR95/01116
; EARLIER FILING DATE: 1995-08-24
; EARLIER APPLICATION NUMBER: FR 94/10299
; EARLIER FILING DATE: 1994-08-25
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 724
; TYPE: PRT
; ORGANISM: B. thuringiensis ser. jegathesan
US-08-793-331-4
```

```
Query Match          46.3%; Score 38; DB 3; Length 724;
Best Local Similarity 70.0%; Pred. No. 2.8e+02;
Matches      7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy      4 FPKLKVEVFP 13
        | ||| |:|
Db      55 FAKLKSEIFP 64
```

Search completed: August 24, 2004, 15:55:20  
Job time : 18.4552 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 24, 2004, 14:54:57 ; Search time 61.1194 Seconds  
(without alignments)  
69.343 Million cell updates/sec

Title: US-09-641-801-8  
Perfect score: 82  
Sequence: 1 LKPFPKLKVEVFPPF 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	82	100.0	15	4	AAB72507	Aab72507 Colostrin
2	82	100.0	15	4	AAB59313	Aab59313 Ewe colos
3	82	100.0	15	4	AAB72253	Aab72253 Colostrin
4	82	100.0	15	4	AAB72539	Aab72539 Colostrin
5	82	100.0	15	5	AAO14584	Aao14584 Neural ce
6	82	100.0	15	5	AAM51043	Aam51043 Colostrin
7	82	100.0	15	5	AAE20235	Aae20235 Colostrin
8	82	100.0	16	4	AAB59344	Aab59344 Ewe colos
9	47	57.3	436	7	ADE71271	Ade71271 Novel hum

10	47	57.3	867	7	ADE71288	Ade71288	Novel hum
11	43	52.4	175	4	ABG16487	Abg16487	Novel hum
12	43	52.4	240	3	AAG57452	Aag57452	Arabidops
13	43	52.4	246	3	AAG57451	Aag57451	Arabidops
14	43	52.4	1007	5	ABB91898	Abb91898	Herbicida
15	42	51.2	116	4	AAO03015	Aao03015	Human pol
16	42	51.2	184	6	ABM69539	Abm69539	Photorhab
17	42	51.2	232	7	ADB64957	Adb64957	Human pro
18	42	51.2	289	5	ABP27516	Abp27516	Streptoco
19	42	51.2	562	4	AAU30319	Aau30319	Novel hum
20	41.5	50.6	73	4	AAU31755	Aau31755	Novel hum
21	41.5	50.6	1675	5	AAU75109	Aau75109	Clathrin
22	41.5	50.6	1675	6	ADA50752	Ada50752	Human cla
23	41	50.0	54	4	AAO04922	Aao04922	Human pol
24	41	50.0	74	4	AAG92276	Aag92276	C glutami
25	41	50.0	77	4	AAM87301	Aam87301	Human imm
26	41	50.0	94	5	ABP03546	Abp03546	Human ORF
27	41	50.0	106	4	AAM06746	Aam06746	Human foe
28	41	50.0	106	4	AAM06758	Aam06758	Human foe
29	41	50.0	106	4	AAM06581	Aam06581	Human foe
30	41	50.0	119	4	AAO04769	Aao04769	Human pol
31	41	50.0	180	4	AAO01266	Aao01266	Human pol
32	41	50.0	188	4	AAU25583	Aau25583	Human G P
33	41	50.0	466	5	ABB93955	Abb93955	Herbicida
34	41	50.0	736	5	ABB55121	Abb55121	Lactococc
35	40.5	49.4	104	4	AAO06528	Aao06528	Human pol
36	40	48.8	65	4	AAO09242	Aao09242	Human pol
37	40	48.8	82	4	ABB17333	Abb17333	Human ner
38	40	48.8	96	4	AAM85760	Aam85760	Human imm
39	40	48.8	121	4	AAO09003	Aao09003	Human pol
40	40	48.8	127	4	AAO04177	Aao04177	Human pol
41	40	48.8	128	4	AAO06127	Aao06127	Human pol
42	40	48.8	152	3	AAG24904	Aag24904	Arabidops
43	40	48.8	152	3	AAG52703	Aag52703	Arabidops
44	40	48.8	336	4	AAB96054	Aab96054	Putative
45	40	48.8	451	2	AAY06314	Aay06314	Pea trunc

# ALIGNMENTS

RESULT 1

AAB72507

ID AAB72507 standard; peptide; 15 AA.

XX

AC AAB72507;

XX

DT 09-MAY-2001 (first entry)

XX

DE Colostrinin peptide #8.

XX

KW Dermatological; oxidative stress regulator; colostrinin.

XX

OS Unidentified.

XX

PN WO200112650-A2.

XX

PD 22-FEB-2001.  
 XX  
 PF 17-AUG-2000; 2000WO-US022665.  
 XX  
 PR 17-AUG-1999; 99US-0149310P.  
 XX  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 XX  
 PI Stanton GJ, Hughes TK, Boldogh I;  
 XX  
 DR WPI; 2001-218342/22.  
 XX  
 PT Modulating oxidative stress level in a cell, involves contacting the cell  
 PT with an oxidative stress regulator selected from colostrinin, its  
 PT constituent peptide, analog or their combinations.  
 XX  
 PS Claim 6; Page 25; 48pp; English.  
 XX  
 CC The present invention relates to a method for modulating the oxidative  
 CC stress level in a cell or a patient, comprising contacting the cell with,  
 CC or administering to the patient, an oxidative stress regulator selected  
 CC from colostrinin, or its constituent peptide (e.g. the present peptide),  
 CC to change the level of an oxidising species in the cell. The method can  
 CC be used to treat oxidative damage to skin, by decreasing or preventing an  
 CC increase in the level of damage to a biomolecule of the patient  
 XX  
 SQ Sequence 15 AA;

Query Match 100.0%; Score 82; DB 4; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LKPFPKLKVEVFPPF 15  
 |||||  
 Db 1 LKPFPKLKVEVFPPF 15

## RESULT 2

AAB59313

ID AAB59313 standard; peptide; 15 AA.

XX

AC AAB59313;

XX

DT 21-MAR-2001 (first entry)

XX

DE Ewe colostrinin peptide fragment A-4.

XX

KW Sheep; colostrinin; proline rich polypeptide; colostrum; immune disorder;

KW central nervous system disorder; dietary supplement; beta-amyloid plaque.

XX

OS Ovis sp.

XX

PN WO200075173-A2.

XX

PD 14-DEC-2000.

XX

PF 02-JUN-2000; 2000WO-GB002128.

XX  
PR 02-JUN-1999; 99GB-00012852.  
XX  
PA (REGE-) REGEN THERAPEUTICS PLC.  
XX  
PI Georgiades JA;  
XX  
DR WPI; 2001-071058/08.  
XX  
PT Peptides having an N-terminal amino acid sequence isolated from  
PT colostrinin for treating e.g. disorders of the central nervous system and  
PT immune system, viral and bacterial infections, and diseases characterized  
PT by amyloid plaques.  
XX  
PS Claim 7; Page 27; 63pp; English.  
XX  
CC The present invention provides the sequences of a number of peptides  
CC found in ewe's colostrinin. Colostrinin is the proline-rich polypeptide  
CC fragment of colostrum. These peptides can be used in the treatment of  
CC central nervous system disorders such as senile dementia, Parkinson's  
CC disease, Alzheimer's disease, psychosis and neurosis, immune system  
CC disorders such as bacterial and viral infections, to improve the  
CC development of a child's immune system, as a dietary supplement, and to  
CC promote the dissolution of beta-amyloid plaques  
XX  
SQ Sequence 15 AA;

Query Match 100.0%; Score 82; DB 4; Length 15;  
Best Local Similarity 100.0%; Pred. No. 4.7e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LKPFPPKLKVEVFPFP 15  
| | | | | | | | | | | | | | |  
Db 1 LKPFPPKLKVEVFPFP 15

RESULT 3

AAB72253

ID AAB72253 standard; peptide; 15 AA.

XX

AC AAB72253;

XX

DT 14-MAY-2001 (first entry)

XX

DE Colostrinin derived cytokine inducing peptide SEQ ID 8.

XX

KW Colostrinin; immune response; cytokine; blood cell proliferation;  
KW central nervous system disorder; neurological disorder; mental disorder;  
KW dementia; neurodegenerative disease; Alzheimer's disease; psychosis;  
KW neurosis; infection.

XX

OS Synthetic.

XX

PN WO200111937-A2.

XX

PD 22-FEB-2001.

XX



PF 17-AUG-2000; 2000WO-US022818.  
 XX  
 PR 17-AUG-1999; 99US-0149311P.  
 XX  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 PA (REGE-) REGEN THERAPEUTICS PLC.  
 XX  
 PI Stanton GJ, Hughes TK, Boldogh I, Georgiades J;  
 XX  
 DR WPI; 2001-202804/20.  
 XX  
 PT Inducing a cytokine and modulating an immune response, useful for  
 PT treating central nervous system diseases and bacterial and viral  
 PT infections, comprises administering colostrinin as an immunological  
 PT regulator.  
 XX  
 PS Claim 1; Page 34; 50pp; English.  
 XX  
 CC Sequences AAB72246 - AAB72275 represent peptides derived from clostrinin,  
 CC a proline rich polypeptide aggregate contained in colostrum. The peptides  
 CC have immune response modulatory activity, and are capable of inducing  
 CC cytokines. Colostrinin and its derived peptides are useful for inducing  
 CC cytokine production, for modulating an immunological response and for  
 CC inducing blood cell proliferation. The peptides are useful in the  
 CC treatment of disorders of the central nervous system, neurological  
 CC disorders, mental disorders, dementia, neurodegenerative diseases,  
 CC Alzheimer's disease, motor neurone disease, psychosis, neurosis, chronic  
 CC disorders of the immune system, bacterial and viral infections and  
 CC acquired immunological deficiencies  
 XX  
 SQ Sequence 15 AA;

Query Match 100.0%; Score 82; DB 4; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LKPFPKLKVEVFPFP 15  
 |||||  
 Db 1 LKPFPKLKVEVFPFP 15

#### RESULT 4

AAB72539

ID AAB72539 standard; peptide; 15 AA.

XX

AC AAB72539;

XX

DT 09-MAY-2001 (first entry)

XX

DE Colostrinin peptide #8.

XX

KW Neuroprotective; neural cell differentiation regulator; colostrinin;  
 KW colostrum.

XX

OS Unidentified.

XX

PN WO200112651-A2.

XX  
 PD 22-FEB-2001.  
 XX  
 PF 17-AUG-2000; 2000WO-US022774.  
 XX  
 PR 17-AUG-1999; 99US-0149633P.  
 XX  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 XX  
 PI Boldogh I;  
 XX  
 DR WPI; 2001-226545/23.  
 XX  
 PT Use of colostrinin, its constituent peptide or analog as a neural cell  
 PT regulator, for promoting neural cell differentiation and treating damaged  
 PT neural cells in a patient.  
 XX  
 PS Claim 6; Page 21; 35pp; English.  
 XX  
 CC The present invention relates to a method for promoting neural cell  
 CC differentiation and treating damaged neural cells, using colostrinin and  
 CC colostrinin constituent peptides (e.g. the present peptide) as a neural  
 CC cell regulator. Colostrinin is a polypeptide complex found in colostrum  
 XX  
 SQ Sequence 15 AA;

Query Match 100.0%; Score 82; DB 4; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LKPFPKLKVEVFPFP 15  
 |||||  
 Db 1 LKPFPKLKVEVFPFP 15

# RESULT 5

AA014584

ID AA014584 standard; peptide; 15 AA.

XX

AC AA014584;

XX

DT 27-MAY-2002 (first entry)

XX

DE Neural cell regulatory colostrinin peptide 8.

XX

KW Neural cell differentiation; neural cell regulator; colostrinin peptide;  
 KW neural cell formation; proline-rich polypeptide aggregate; colostrum;  
 KW neural cell treatment.

XX

OS Unidentified.

XX

FH Key Location/Qualifiers

FT Modified-site 15

FT /note= "Optional C-terminal amide"

XX

PN WO200213851-A1.

XX

PD 21-FEB-2002.  
 XX  
 PF 17-AUG-2000; 2000WO-US022777.  
 XX  
 PR 17-AUG-2000; 2000WO-US022777.  
 XX  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 XX  
 PI Boldogh I, Stanton JG, Hughes TK;  
 XX  
 DR WPI; 2002-269152/31.  
 XX  
 PT Promoting cell differentiation in a patient involves use of blood cell  
 PT regulator selected from colostrinin, its constituent peptide and/or  
 PT analog.  
 XX  
 PS Claim 7; Page 21; 37pp; English.  
 XX  
 CC The invention comprises a method for promoting cell differentiation (e.g.  
 CC neural cell differentiation). The method involves contacting cells with a  
 CC neural cell regulator (i.e. a colostrinin peptide) in order to change the  
 CC cells in morphology to form neural cells. Colostrinin is a proline-rich  
 CC polypeptide aggregate that is present in colostrum. The method of the  
 CC invention is useful for promoting the differentiation of cells and for  
 CC treating damaged neural cells in a patient. The present amino acid  
 CC sequence represents a specifically claimed colostrinin peptide used in  
 CC the method of the invention  
 XX  
 SQ Sequence 15 AA;

Query Match 100.0%; Score 82; DB 5; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LKPFPKLKVEVFPFP 15  
 |||||  
 Db 1 LKPFPKLKVEVFPFP 15

# RESULT 6

AAM51043

ID AAM51043 standard; peptide; 15 AA.

XX

AC AAM51043;

XX

DT 30-MAY-2002 (first entry)

XX

DE Colostrinin constituent peptide.

XX

KW Colostrinin; colostrum; immunomodulator; cardiovascular;

KW blood cell regulator; cytokine inducer; human.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Modified-site 15

FT /note= "optional C-terminal amidation"

XX  
 PN WO200213849-A1.  
 XX  
 PD 21-FEB-2002.  
 XX  
 PF 17-AUG-2000; 2000WO-US022775.  
 XX  
 PR 17-AUG-2000; 2000WO-US022775.  
 XX  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 PA (REGE-) REGEN THERAPEUTICS PLC.  
 XX  
 PI Stanton GJ, Hughes TK, Boldogh I, Georgiades J;  
 XX  
 DR WPI; 2002-269150/31.  
 XX  
 PT Modulation of blood cell proliferation in a patient involves use of blood  
 PT cell regulator selected from colostrinin, its constituent peptide and/or  
 PT analog.  
 XX  
 PS Claim 1; Page 34; 54pp; English.  
 XX  
 CC The present sequence is that of a colostrinin constituent peptide that is  
 CC preferred for use as an immunological regulator and as a blood cell  
 CC regulator in claimed methods of the invention. Methods are claimed for:  
 CC inducing a cytokine in a cell by contact with an immunological regulator,  
 CC where the cell is present in a cell culture, a tissue, an organ or an  
 CC organism, and the cell is mammalian, including human; modulating an  
 CC immune response in a cell by contact with the immunological regulator  
 CC under conditions effective to induce a cytokine; modulating an immune  
 CC response in a patient by administering an immunological regulator under  
 CC conditions effective to induce a cytokine, where the immunological  
 CC regulator is administered topically or as part of a dietary supplement,  
 CC and where the immune response is specific or non specific, an interferon  
 CC response or an antibody response; modulating blood cell proliferation by  
 CC contacting blood cells with a blood cell regulator, where the blood cells  
 CC are present in a cell culture or an organism, are mammalian or human, and  
 CC where the blood cells are increased in number or differentiated; and a  
 CC method for modulating blood cell proliferation in a patent. A claimed  
 CC cytokine-inducing composition comprises a pharmaceutical carrier and an  
 CC active agent such as the present peptide. Cytokines induced by this  
 CC peptide in human leucocyte cultures include interferon-gamma, tumour  
 CC necrosis factor-alpha, interleukin-6 and interleukin-10  
 XX  
 SQ Sequence 15 AA;

Query Match 100.0%; Score 82; DB 5; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LKPFPPKLKVEVFPFP 15  
 |||||  
 Db 1 LKPFPPKLKVEVFPFP 15

RESULT 7  
 AAE20235

ID AAE20235 standard; peptide; 15 AA.  
XX  
AC AAE20235;  
XX  
DT 18-JUN-2002 (first entry)  
XX  
DE Colostrinin constituent peptide #8.  
XX  
KW Blood cell regulator; colostrinin; constituent peptide; oxidative stress;  
KW therapy; oxidative damage; skin; aging; wound healing; cell replacement;  
KW tissue; organ; cosmetic procedure; repair; regeneration; preservation;  
KW transplantation; implantation; dermatological; vulnerary.  
XX  
OS Unidentified.  
XX  
FH Key Location/Qualifiers  
FT Modified-site 15  
FT /note= "Optionally C-terminal amide"  
XX  
PN WO200213850-A1.  
XX  
PD 21-FEB-2002.  
XX  
PF 17-AUG-2000; 2000WO-US022776.  
XX  
PR 17-AUG-2000; 2000WO-US022776.  
XX  
PA (TEXA ) UNIV TEXAS SYSTEM.  
XX  
PI Stanton GJ, Hughes TK, Boldogh I;  
XX  
DR WPI; 2002-269151/31.  
XX  
PT Composition useful for the modulation of blood cell proliferation in a  
PT patient comprises a blood cell regulator selected from colostrinin, its  
PT constituent peptide and/or analog.  
XX  
PS Claim 6; Page 25; 51pp; English.  
XX  
CC The invention relates to a composition which comprises a blood cell  
CC regulator selected from colostrinin, its constituent peptide and/or  
CC analogue. The invention is used for modulating the oxidative stress level  
CC in a cell e.g. mammalian or human cell present in a cell culture, tissue,  
CC organ, or organism; or for treating oxidative damage to the skin of a  
CC patient e.g. animal or human; to modulate oxidative stress during/ after  
CC a premature birth or normal birth, preventing/delaying aging in a  
CC patient, enhancing wound healing, and the reduction of side effects of  
CC cosmetic procedures. The method changes the level of an oxidising species  
CC in the cell, such as decreases or prevents increase in the level of  
CC damage to a biomolecule of the patient selected from DNA, protein and/or  
CC lipid, compared to the same conditions when the oxidative stress  
CC regulator is not present. The modulation of oxidative stress results in  
CC enhanced repair, regeneration, and replacement of cells, tissues and  
CC organs (e.g. kidney, liver, pancreas, skin, and the other internal and  
CC external organs), as well as enhanced preservation of such organs for  
CC transplantation, implantation, or scientific research. The present  
CC sequence is a colostrinin constituent peptide

XX

SQ Sequence 15 AA;

Query Match 100.0%; Score 82; DB 5; Length 15;  
Best Local Similarity 100.0%; Pred. No. 4.7e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LKPFPPKLKVEVFPPF 15  
| | | | | | | | | | | | | | |  
Db 1 LKPFPPKLKVEVFPPF 15

RESULT 8

AAB59344

ID AAB59344 standard; peptide; 16 AA.

XX

AC AAB59344;

XX

DT 21-MAR-2001 (first entry)

XX

DE Ewe colostrinin peptide fragment derived sequence #4.

XX

KW Sheep; colostrinin; proline rich polypeptide; colostrum; immune disorder;  
KW central nervous system disorder; dietary supplement; beta-amyloid plaque.

XX

OS Ovis sp.

XX

PN WO200075173-A2.

XX

PD 14-DEC-2000.

XX

PF 02-JUN-2000; 2000WO-GB002128.

XX

PR 02-JUN-1999; 99GB-00012852.

XX

PA (REGE-) REGEN THERAPEUTICS PLC.

XX

PI Georgiades JA;

XX

DR WPI; 2001-071058/08.

XX

PT Peptides having an N-terminal amino acid sequence isolated from  
PT colostrinin for treating e.g. disorders of the central nervous system and  
PT immune system, viral and bacterial infections, and diseases characterized  
PT by amyloid plaques.

XX

PS Claim 8; Page 27; 63pp; English.

XX

CC The present invention provides the sequences of a number of peptides  
CC found in ewe's colostrinin. Colostrinin is the proline-rich polypeptide  
CC fragment of colostrum. These peptides can be used in the treatment of  
CC central nervous system disorders such as senile dementia, Parkinson's  
CC disease, Alzheimer's disease, psychosis and neurosis, immune system  
CC disorders such as bacterial and viral infections, to improve the  
CC development of a child's immune system, as a dietary supplement, and to  
CC promote the dissolution of beta-amyloid plaques

XX

SQ Sequence 16 AA;

Query Match 100.0%; Score 82; DB 4; Length 16;  
Best Local Similarity 100.0%; Pred. No. 5.1e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LKPFPKLKVEVFPP 15  
| | | | | | | | | | | | | |  
Db 2 LKPFPKLKVEVFPP 16

RESULT 9

ADE71271

ID ADE71271 standard; protein; 436 AA.

XX

AC ADE71271;

XX

DT 29-JAN-2004 (first entry)

XX

DE Novel human protein #25.

XX

KW human; novel protein; drug.

XX

OS Homo sapiens.

XX

PN JP2002345493-A.

XX

PD 03-DEC-2002.

XX

PF 29-MAR-2001; 2002JP-00049046.

XX

PR 29-MAR-2001; 2001JP-00095524.

XX

PA (KAZU-) ZH KAZUSA DNA KENKYUSHO.

XX

DR WPI; 2003-460885/44.

DR N-PSDB; ADE71209.

XX

PT A gene and a protein encoded by it, used in drugs.

XX

PS Disclosure; Page 126-128; 257pp; Japanese.

XX

CC The invention comprises the amino acid and coding sequences of novel  
CC human proteins. The DNA and protein sequences of the invention are used  
CC in drugs. The present amino acid sequence represents a novel human  
CC protein of the invention.

XX

SQ Sequence 436 AA;

Query Match 57.3%; Score 47; DB 7; Length 436;  
Best Local Similarity 61.5%; Pred. No. 15;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 LKPFPKLKVEVF 13  
| | : | | | : | |  
Db 358 LTPYPKLKLTALFP 370

RESULT 10

ADE71288

ID ADE71288 standard; protein; 867 AA.

XX

AC ADE71288;

XX

DT 29-JAN-2004 (first entry)

XX

DE Novel human protein #42.

XX

KW human; novel protein; drug.

XX

OS Homo sapiens.

XX

PN JP2002345493-A.

XX

PD 03-DEC-2002.

XX

PF 29-MAR-2001; 2002JP-00049046.

XX

PR 29-MAR-2001; 2001JP-00095524.

XX

PA (KAZU-) ZH KAZUSA DNA KENKYUSHO.

XX

DR WPI; 2003-460885/44.

DR N-PSDB; ADE71226.

XX

PT A gene and a protein encoded by it, used in drugs.

XX

PS Disclosure; Page 189-192; 257pp; Japanese.

XX

CC The invention comprises the amino acid and coding sequences of novel  
 CC human proteins. The DNA and protein sequences of the invention are used  
 CC in drugs. The present amino acid sequence represents a novel human  
 CC protein of the invention.

XX

SQ Sequence 867 AA;

Query Match 57.3%; Score 47; DB 7; Length 867;

Best Local Similarity 61.5%; Pred. No. 32;

Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 LKPFPKLKVEVFP 13

| | : | | | : | |

Db 114 LTPYPKLTALFP 126

RESULT 11

ABG16487

ID ABG16487 standard; protein; 175 AA.

XX

AC ABG16487;

XX

DT 18-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #16478.



XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US008631.  
 XX  
 PR 31-MAR-2000; 2000US-00540217.  
 PR 23-AUG-2000; 2000US-00649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS80674.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX  
 PS Claim 20; SEQ ID NO 46846; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 175 AA;

Query Match 52.4%; Score 43; DB 4; Length 175;  
 Best Local Similarity 61.5%; Pred. No. 27;  
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy            1 LKPFPKLKVEVFP 13  
              | ||| || :||  
Db           27 LPPFPPLKFFIFP 39

RESULT 12

AAG57452

ID    AAG57452 standard; protein; 240 AA.

XX

AC    AAG57452;

XX

DT    18-OCT-2000    (first entry)

XX

DE    Arabidopsis thaliana protein fragment SEQ ID NO: 74041.

XX

KW    Protein identification; signal transduction pathway; metabolic pathway;  
KW    hybridisation assay; genetic mapping; gene expression control; promoter;  
KW    termination sequence.

XX

OS    Arabidopsis thaliana.

XX

PN    EP1033405-A2.

XX

PD    06-SEP-2000.

XX

PF    25-FEB-2000; 2000EP-00301439.

XX

PR    25-FEB-1999;    99US-0121825P.

PR    05-MAR-1999;    99US-0123180P.

PR    09-MAR-1999;    99US-0123548P.

PR    23-MAR-1999;    99US-0125788P.

PR    25-MAR-1999;    99US-0126264P.

PR    29-MAR-1999;    99US-0126785P.

PR    01-APR-1999;    99US-0127462P.

PR    06-APR-1999;    99US-0128234P.

PR    08-APR-1999;    99US-0128714P.

PR    16-APR-1999;    99US-0129845P.

PR    19-APR-1999;    99US-0130077P.

PR    21-APR-1999;    99US-0130449P.

PR    23-APR-1999;    99US-0130510P.

PR    23-APR-1999;    99US-0130891P.

PR    28-APR-1999;    99US-0131449P.

PR    30-APR-1999;    99US-0132048P.

PR    30-APR-1999;    99US-0132407P.

PR    04-MAY-1999;    99US-0132484P.

PR    05-MAY-1999;    99US-0132485P.

PR    06-MAY-1999;    99US-0132486P.

PR    06-MAY-1999;    99US-0132487P.

PR    07-MAY-1999;    99US-0132863P.

PR    11-MAY-1999;    99US-0134256P.

PR    14-MAY-1999;    99US-0134218P.

PR    14-MAY-1999;    99US-0134219P.

PR    14-MAY-1999;    99US-0134221P.

PR    14-MAY-1999;    99US-0134370P.

PR    18-MAY-1999;    99US-0134768P.

PR    19-MAY-1999;    99US-0134941P.

PR    20-MAY-1999;    99US-0135124P.

PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.

PR	21-JUL-1999;	99US-0144814P.
PR	21-JUL-1999;	99US-0145086P.
PR	21-JUL-1999;	99US-0145088P.
PR	22-JUL-1999;	99US-0145085P.
PR	22-JUL-1999;	99US-0145087P.
PR	22-JUL-1999;	99US-0145089P.
PR	22-JUL-1999;	99US-0145192P.
PR	23-JUL-1999;	99US-0145145P.
PR	23-JUL-1999;	99US-0145218P.
PR	23-JUL-1999;	99US-0145224P.
PR	26-JUL-1999;	99US-0145276P.
PR	27-JUL-1999;	99US-0145913P.
PR	27-JUL-1999;	99US-0145918P.
PR	27-JUL-1999;	99US-0145919P.
PR	28-JUL-1999;	99US-0145951P.
PR	02-AUG-1999;	99US-0146386P.
PR	02-AUG-1999;	99US-0146388P.
PR	02-AUG-1999;	99US-0146389P.
PR	03-AUG-1999;	99US-0147038P.
PR	04-AUG-1999;	99US-0147204P.
PR	04-AUG-1999;	99US-0147302P.
PR	05-AUG-1999;	99US-0147192P.
PR	05-AUG-1999;	99US-0147260P.
PR	06-AUG-1999;	99US-0147303P.
PR	06-AUG-1999;	99US-0147416P.
PR	09-AUG-1999;	99US-0147493P.
PR	09-AUG-1999;	99US-0147935P.
PR	10-AUG-1999;	99US-0148171P.
PR	11-AUG-1999;	99US-0148319P.
PR	12-AUG-1999;	99US-0148341P.
PR	13-AUG-1999;	99US-0148565P.
PR	13-AUG-1999;	99US-0148684P.
PR	16-AUG-1999;	99US-0149368P.
PR	17-AUG-1999;	99US-0149175P.
PR	18-AUG-1999;	99US-0149426P.
PR	20-AUG-1999;	99US-0149722P.
PR	20-AUG-1999;	99US-0149723P.
PR	20-AUG-1999;	99US-0149929P.
PR	23-AUG-1999;	99US-0149902P.
PR	23-AUG-1999;	99US-0149930P.
PR	25-AUG-1999;	99US-0150566P.
PR	26-AUG-1999;	99US-0150884P.
PR	27-AUG-1999;	99US-0151065P.
PR	27-AUG-1999;	99US-0151066P.
PR	27-AUG-1999;	99US-0151080P.
PR	30-AUG-1999;	99US-0151303P.
PR	31-AUG-1999;	99US-0151438P.
PR	01-SEP-1999;	99US-0151930P.
PR	07-SEP-1999;	99US-0152363P.
PR	10-SEP-1999;	99US-0153070P.
PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155139P.
PR	23-SEP-1999;	99US-0155486P.
PR	24-SEP-1999;	99US-0155659P.

PR 28-SEP-1999; 99US-0156458P.  
 PR 29-SEP-1999; 99US-0156596P.  
 PR 04-OCT-1999; 99US-0157117P.  
 PR 05-OCT-1999; 99US-0157753P.  
 PR 06-OCT-1999; 99US-0157865P.  
 PR 07-OCT-1999; 99US-0158029P.  
 PR 08-OCT-1999; 99US-0158232P.  
 PR 12-OCT-1999; 99US-0158369P.  
 PR 13-OCT-1999; 99US-0159293P.  
 PR 13-OCT-1999; 99US-0159294P.  
 PR 13-OCT-1999; 99US-0159295P.  
 PR 14-OCT-1999; 99US-0159329P.  
 PR 14-OCT-1999; 99US-0159330P.  
 PR 14-OCT-1999; 99US-0159331P.  
 PR 14-OCT-1999; 99US-0159637P.  
 PR 14-OCT-1999; 99US-0159638P.  
 PR 18-OCT-1999; 99US-0159584P.  
 PR 21-OCT-1999; 99US-0160741P.  
 PR 21-OCT-1999; 99US-0160767P.  
 PR 21-OCT-1999; 99US-0160768P.  
 PR 21-OCT-1999; 99US-0160770P.  
 PR 21-OCT-1999; 99US-0160814P.  
 PR 21-OCT-1999; 99US-0160815P.  
 PR 22-OCT-1999; 99US-0160980P.  
 PR 22-OCT-1999; 99US-0160981P.  
 PR 22-OCT-1999; 99US-0160989P.  
 PR 25-OCT-1999; 99US-0161404P.  
 PR 25-OCT-1999; 99US-0161405P.  
 PR 25-OCT-1999; 99US-0161406P.  
 PR 26-OCT-1999; 99US-0161359P.  
 PR 26-OCT-1999; 99US-0161360P.  
 PR 26-OCT-1999; 99US-0161361P.  
 PR 28-OCT-1999; 99US-0161920P.  
 PR 28-OCT-1999; 99US-0161992P.  
 PR 28-OCT-1999; 99US-0161993P.  
 PR 29-OCT-1999; 99US-0162142P.

Query Match 52.4%; Score 43; DB 3; Length 240;  
 Best Local Similarity 61.5%; Pred. No. 38;  
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 3 PFPKLVVEVFPFP 15  
 | | | ||: |||  
 Db 203 PTPHLVVEITPFP 215

# RESULT 13

AAG57451

ID AAG57451 standard; protein; 246 AA.

XX

AC AAG57451;

XX

DT 18-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 74040.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 XX  
 OS *Arabidopsis thaliana*.  
 XX  
 PN EP1033405-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 25-FEB-2000; 2000EP-00301439.  
 XX  
 PR 25-FEB-1999; 99US-0121825P.  
 PR 05-MAR-1999; 99US-0123180P.  
 PR 09-MAR-1999; 99US-0123548P.  
 PR 23-MAR-1999; 99US-0125788P.  
 PR 25-MAR-1999; 99US-0126264P.  
 PR 29-MAR-1999; 99US-0126785P.  
 PR 01-APR-1999; 99US-0127462P.  
 PR 06-APR-1999; 99US-0128234P.  
 PR 08-APR-1999; 99US-0128714P.  
 PR 16-APR-1999; 99US-0129845P.  
 PR 19-APR-1999; 99US-0130077P.  
 PR 21-APR-1999; 99US-0130449P.  
 PR 23-APR-1999; 99US-0130510P.  
 PR 23-APR-1999; 99US-0130891P.  
 PR 28-APR-1999; 99US-0131449P.  
 PR 30-APR-1999; 99US-0132048P.  
 PR 30-APR-1999; 99US-0132407P.  
 PR 04-MAY-1999; 99US-0132484P.  
 PR 05-MAY-1999; 99US-0132485P.  
 PR 06-MAY-1999; 99US-0132486P.  
 PR 06-MAY-1999; 99US-0132487P.  
 PR 07-MAY-1999; 99US-0132863P.  
 PR 11-MAY-1999; 99US-0134256P.  
 PR 14-MAY-1999; 99US-0134218P.  
 PR 14-MAY-1999; 99US-0134219P.  
 PR 14-MAY-1999; 99US-0134221P.  
 PR 14-MAY-1999; 99US-0134370P.  
 PR 18-MAY-1999; 99US-0134768P.  
 PR 19-MAY-1999; 99US-0134941P.  
 PR 20-MAY-1999; 99US-0135124P.  
 PR 21-MAY-1999; 99US-0135353P.  
 PR 24-MAY-1999; 99US-0135629P.  
 PR 25-MAY-1999; 99US-0136021P.  
 PR 27-MAY-1999; 99US-0136392P.  
 PR 28-MAY-1999; 99US-0136782P.  
 PR 01-JUN-1999; 99US-0137222P.  
 PR 03-JUN-1999; 99US-0137528P.  
 PR 04-JUN-1999; 99US-0137502P.  
 PR 07-JUN-1999; 99US-0137724P.  
 PR 08-JUN-1999; 99US-0138094P.  
 PR 10-JUN-1999; 99US-0138540P.  
 PR 10-JUN-1999; 99US-0138847P.  
 PR 14-JUN-1999; 99US-0139119P.  
 PR 16-JUN-1999; 99US-0139452P.  
 PR 16-JUN-1999; 99US-0139453P.  
 PR 17-JUN-1999; 99US-0139492P.

PR	18-JUN-1999;	99US-0139454P.
PR	18-JUN-1999;	99US-0139455P.
PR	18-JUN-1999;	99US-0139456P.
PR	18-JUN-1999;	99US-0139457P.
PR	18-JUN-1999;	99US-0139458P.
PR	18-JUN-1999;	99US-0139459P.
PR	18-JUN-1999;	99US-0139460P.
PR	18-JUN-1999;	99US-0139461P.
PR	18-JUN-1999;	99US-0139462P.
PR	18-JUN-1999;	99US-0139463P.
PR	18-JUN-1999;	99US-0139750P.
PR	18-JUN-1999;	99US-0139763P.
PR	21-JUN-1999;	99US-0139817P.
PR	22-JUN-1999;	99US-0139899P.
PR	23-JUN-1999;	99US-0140353P.
PR	23-JUN-1999;	99US-0140354P.
PR	24-JUN-1999;	99US-0140695P.
PR	28-JUN-1999;	99US-0140823P.
PR	29-JUN-1999;	99US-0140991P.
PR	30-JUN-1999;	99US-0141287P.
PR	01-JUL-1999;	99US-0141842P.
PR	01-JUL-1999;	99US-0142154P.
PR	02-JUL-1999;	99US-0142055P.
PR	06-JUL-1999;	99US-0142390P.
PR	08-JUL-1999;	99US-0142803P.
PR	09-JUL-1999;	99US-0142920P.
PR	12-JUL-1999;	99US-0142977P.
PR	13-JUL-1999;	99US-0143542P.
PR	14-JUL-1999;	99US-0143624P.
PR	15-JUL-1999;	99US-0144005P.
PR	16-JUL-1999;	99US-0144085P.
PR	16-JUL-1999;	99US-0144086P.
PR	19-JUL-1999;	99US-0144325P.
PR	19-JUL-1999;	99US-0144331P.
PR	19-JUL-1999;	99US-0144332P.
PR	19-JUL-1999;	99US-0144333P.
PR	19-JUL-1999;	99US-0144334P.
PR	19-JUL-1999;	99US-0144335P.
PR	20-JUL-1999;	99US-0144352P.
PR	20-JUL-1999;	99US-0144632P.
PR	20-JUL-1999;	99US-0144884P.
PR	21-JUL-1999;	99US-0144814P.
PR	21-JUL-1999;	99US-0145086P.
PR	21-JUL-1999;	99US-0145088P.
PR	22-JUL-1999;	99US-0145085P.
PR	22-JUL-1999;	99US-0145087P.
PR	22-JUL-1999;	99US-0145089P.
PR	22-JUL-1999;	99US-0145192P.
PR	23-JUL-1999;	99US-0145145P.
PR	23-JUL-1999;	99US-0145218P.
PR	23-JUL-1999;	99US-0145224P.
PR	26-JUL-1999;	99US-0145276P.
PR	27-JUL-1999;	99US-0145913P.
PR	27-JUL-1999;	99US-0145918P.
PR	27-JUL-1999;	99US-0145919P.
PR	28-JUL-1999;	99US-0145951P.
PR	02-AUG-1999;	99US-0146386P.

PR	02-AUG-1999;	99US-0146388P.
PR	02-AUG-1999;	99US-0146389P.
PR	03-AUG-1999;	99US-0147038P.
PR	04-AUG-1999;	99US-0147204P.
PR	04-AUG-1999;	99US-0147302P.
PR	05-AUG-1999;	99US-0147192P.
PR	05-AUG-1999;	99US-0147260P.
PR	06-AUG-1999;	99US-0147303P.
PR	06-AUG-1999;	99US-0147416P.
PR	09-AUG-1999;	99US-0147493P.
PR	09-AUG-1999;	99US-0147935P.
PR	10-AUG-1999;	99US-0148171P.
PR	11-AUG-1999;	99US-0148319P.
PR	12-AUG-1999;	99US-0148341P.
PR	13-AUG-1999;	99US-0148565P.
PR	13-AUG-1999;	99US-0148684P.
PR	16-AUG-1999;	99US-0149368P.
PR	17-AUG-1999;	99US-0149175P.
PR	18-AUG-1999;	99US-0149426P.
PR	20-AUG-1999;	99US-0149722P.
PR	20-AUG-1999;	99US-0149723P.
PR	20-AUG-1999;	99US-0149929P.
PR	23-AUG-1999;	99US-0149902P.
PR	23-AUG-1999;	99US-0149930P.
PR	25-AUG-1999;	99US-0150566P.
PR	26-AUG-1999;	99US-0150884P.
PR	27-AUG-1999;	99US-0151065P.
PR	27-AUG-1999;	99US-0151066P.
PR	27-AUG-1999;	99US-0151080P.
PR	30-AUG-1999;	99US-0151303P.
PR	31-AUG-1999;	99US-0151438P.
PR	01-SEP-1999;	99US-0151930P.
PR	07-SEP-1999;	99US-0152363P.
PR	10-SEP-1999;	99US-0153070P.
PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155139P.
PR	23-SEP-1999;	99US-0155486P.
PR	24-SEP-1999;	99US-0155659P.
PR	28-SEP-1999;	99US-0156458P.
PR	29-SEP-1999;	99US-0156596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157753P.
PR	06-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158232P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0159293P.
PR	13-OCT-1999;	99US-0159294P.
PR	13-OCT-1999;	99US-0159295P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	14-OCT-1999;	99US-0159638P.



PR 18-OCT-1999; 99US-0159584P.  
 PR 21-OCT-1999; 99US-0160741P.  
 PR 21-OCT-1999; 99US-0160767P.  
 PR 21-OCT-1999; 99US-0160768P.  
 PR 21-OCT-1999; 99US-0160770P.  
 PR 21-OCT-1999; 99US-0160814P.  
 PR 21-OCT-1999; 99US-0160815P.  
 PR 22-OCT-1999; 99US-0160980P.  
 PR 22-OCT-1999; 99US-0160981P.  
 PR 22-OCT-1999; 99US-0160989P.  
 PR 25-OCT-1999; 99US-0161404P.  
 PR 25-OCT-1999; 99US-0161405P.  
 PR 25-OCT-1999; 99US-0161406P.  
 PR 26-OCT-1999; 99US-0161359P.  
 PR 26-OCT-1999; 99US-0161360P.  
 PR 26-OCT-1999; 99US-0161361P.  
 PR 28-OCT-1999; 99US-0161920P.  
 PR 28-OCT-1999; 99US-0161992P.  
 PR 28-OCT-1999; 99US-0161993P.  
 PR 29-OCT-1999; 99US-0162142P.

Query Match 52.4%; Score 43; DB 3; Length 246;  
 Best Local Similarity 61.5%; Pred. No. 39;  
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 3 PFPKLVVEVFPFP 15  
 | | | ||: |||  
 Db 209 PTPHLVVEITPFP 221

# RESULT 14

ABB91898

ID ABB91898 standard; protein; 1007 AA.

XX

AC ABB91898;

XX

DT 31-MAY-2002 (first entry)

XX

DE Herbicidally active polypeptide SEQ ID NO 1109.

XX

KW Herbicidal; plant; agriculture; herbicide.

XX

OS Arabidopsis thaliana.

XX

PN WO200210210-A2.

XX

PD 07-FEB-2002.

XX

PF 28-AUG-2001; 2001WO-EP009892.

XX

PR 28-AUG-2001; 2001WO-EP009892.

XX

PA (FARB ) BAYER AG.

XX

PI Tietjen K, Weidler M;

XX

DR WPI; 2002-269010/31.

XX  
PT Identifying plant target proteins for herbicidally active compounds,  
PT comprising aligning and comparing nucleic acid or amino acid sequences  
PT from plant with nucleic acid or amino acid sequences from non-plant  
PT organisms.  
XX  
PS Claim 5; SEQ ID NO 1109; 261pp + Sequence Listing; English.  
XX  
CC The invention relates to identifying target proteins (ABB90790-ABB94016)  
CC for herbicidally active compounds, comprising aligning and comparing  
CC nucleic acid or amino acid sequences from plant with nucleic acid or  
CC amino acid sequences from non-plant organisms using suitable search  
CC parameters, where plant sequences having an E-value greater by a factor  
CC of 3 than the E-value of most similar non-plant sequences are selected.  
CC The polypeptides or nucleic acids encoding them are useful for  
CC identifying modulators. The identified modulators are useful as  
CC herbicides  
XX  
SQ Sequence 1007 AA;

Query Match 52.4%; Score 43; DB 5; Length 1007;  
Best Local Similarity 53.8%; Pred. No. 1.8e+02;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 2 KPFPKCLKVEVFPPF 14  
||:| || :|: |  
Db 898 KPYPCLKSDVYAF 910

# RESULT 15

AAO03015

ID AAO03015 standard; protein; 116 AA.

XX

AC AAO03015;

XX

DT 06-NOV-2001 (first entry)

XX

DE Human polypeptide SEQ ID NO 16907.

XX

KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.

XX

OS Homo sapiens.

XX

PN WO200164835-A2.

XX

PD 07-SEP-2001.

XX

PF 26-FEB-2001; 2001WO-US004927.

XX

PR 28-FEB-2000; 2000US-00515126.

PR 18-MAY-2000; 2000US-00577409.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Tang YT, Liu C, Drmanac RT;

XX

DR WPI; 2001-514838/56.

DR N-PSDB; AAI82946.

XX

PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing  
PT and treating e.g. leukemia, inflammation and immune disorders.

XX

PS Claim 20; SEQ ID NO 16907; 1399pp + Sequence Listing; English.

XX

CC The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation. Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 116 AA;

Query Match 51.2%; Score 42; DB 4; Length 116;

Best Local Similarity 61.5%; Pred. No. 25;

Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 PFPKLVKVEVFPP 15

||| || || |

Db 29 PFPPLKFXXFFPKP 41

Search completed: August 24, 2004, 15:42:41

Job time : 67.1194 secs

OM protein - protein search, using sw model

Run on: August 24, 2004, 15:26:28 ; Search time 14.5522 Seconds  
(without alignments)  
99.151 Million cell updates/sec

Title: US-09-641-801-8  
Perfect score: 82  
Sequence: 1 LKPFPKLKVEVFPFP 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	43	52.4	127	2	F86175	protein F19P19.17
2	43	52.4	790	2	G90477	hypothetical prote
3	43	52.4	1007	2	C84668	probable receptor-
4	42	51.2	190	2	T45013	hypothetical prote
5	42	51.2	343	2	T28908	hypothetical prote
6	42	51.2	428	2	AD2650	oxidoreductase ord
7	42	51.2	428	2	C97432	probable oxidoredu
8	42	51.2	939	2	T18974	hypothetical prote
9	42	51.2	965	2	T22933	hypothetical prote
10	41.5	50.6	379	2	S44745	C02D5.3 protein -
11	41.5	50.6	1675	1	LRRTH	clathrin heavy cha
12	41	50.0	398	2	T38233	probable cystathio
13	41	50.0	401	2	A82221	extracellular solu

14	41	50.0	496	2	S75790	hypothetical prote
15	41	50.0	537	2	C64432	hypothetical prote
16	41	50.0	736	2	D86846	competence protein
17	41	50.0	1353	2	T19157	probable metal bin
18	41	50.0	1424	2	T19156	probable metal bin
19	40.5	49.4	1556	2	S59393	probable membrane
20	40	48.8	134	2	T32913	hypothetical prote
21	40	48.8	134	2	C87723	protein C54G6.4 [i
22	40	48.8	193	2	E86233	hypothetical prote
23	40	48.8	212	2	H98291	hypothetical prote
24	40	48.8	212	2	AB2992	hypothetical prote
25	40	48.8	259	1	YXBYT	thymidylate syntha
26	40	48.8	336	2	B75188	probable 4-hydroxy
27	40	48.8	424	2	T33839	hypothetical prote
28	40	48.8	572	2	T34345	hypothetical prote
29	39	47.6	258	2	T51898	hypothetical prote
30	39	47.6	307	1	YXMST	thymidylate syntha
31	39	47.6	327	2	JS0402	gamma-gliadin prec
32	39	47.6	371	2	D97042	hypothetical prote
33	39	47.6	385	2	G90274	hypothetical prote
34	39	47.6	481	2	A98947	probable rRNA meth
35	39	47.6	481	2	E85795	probable nucleolar
36	39	47.6	481	2	C64945	probable rRNA meth
37	39	47.6	509	2	T50006	receptor protein k
38	39	47.6	562	2	T05758	hypothetical prote
39	39	47.6	681	2	AF0697	probable type III
40	39	47.6	706	2	G71004	hypothetical prote
41	39	47.6	782	2	A82940	hypothetical prote
42	39	47.6	1263	2	AH2011	heterocyst glycoli
43	39	47.6	1791	2	T02345	hypothetical prote
44	38.5	47.0	270	2	AD1408	hypothetical prote
45	38.5	47.0	270	2	AD1784	hypothetical prote

## ALIGNMENTS

### RESULT 1

F86175

protein F19P19.17 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C;Accession: F86175

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Altaf, H.; Araujo, R.; Bowman, C.L.; Brooks, S.Y.; Buehler, E.; Chan, A.; Chao, Q.; Chen, H.; Cheuk, R.F.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Dunn, P.; Etgu, P.; Feldblyum, T.V.; Feng, J.; Fong, B.; Fujii, C.Y.; Gill, J.E.; Goldsmith, A.D.; Haas, B.; Hansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.J.; Koo, H.L.; Kremenetskaia, I.; Kurtz, D.B.; Kwan, A.; Lam, B.; Langin-Hooper, S.; Lee, A.; Lee, J.M.; Lenz, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, A.; Militscher, J.; Miranda, M.; Nguyen, M.; Nierman, W.C.; Osborne, B.I.; Pai, G.; Peterson, J.; Pham, P.K.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, L.J.; Tambunga, G.; Toriumi, M.J.; Town, C.D.; Utterback, T.; van Aken, S.; Vaysberg, M.; Vysotskaia, V.S.; Walker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A;Reference number: A86141; MUID:21016719; PMID:11130712  
A;Accession: F86175  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-127 <STO>  
A;Cross-references: GB:AE005172; NID:g2341037; PIDN:AAB70437.1; GSPDB:GN00141  
C;Genetics:  
A;Gene: F19P19.17  
A;Map position: 1

Query Match 52.4%; Score 43; DB 2; Length 127;  
Best Local Similarity 80.0%; Pred. No. 4.3;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LKPFPKLKVE 10  
||||:|||  
Db 59 LKPFPRKSE 68

## RESULT 2

G90477

hypothetical protein soxM [imported] - Sulfolobus solfataricus

C;Species: Sulfolobus solfataricus

C;Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 03-Aug-2001

C;Accession: G90477

R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-Weiher, C.C.Y.; Clausen, I.G.; Curtis, B.A.; De Moors, A.; Erauso, G.; Fletcher, C.; Gordon, P.M.K.; Heikamp-de Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.; Schenk, M.E.; Theriault, C.; Tolstrup, N.; Charlebois, R.L.; Doolittle, W.F.; Duguet, M.; Gaasterland, T.; Garrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.  
submitted to GenBank, April 2001

A;Description: Sulfolobus solfataricus complete genome.

A;Reference number: A99139

A;Accession: G90477

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-790 <KUR>

A;Cross-references: GB:AE006641; NID:g13816356; PIDN:AAK43078.1; GSPDB:GN00155

C;Genetics:

A;Gene: soxM

C;Superfamily: cytochrome-c oxidase chain I/III; cytochrome-c oxidase chain I homology

C;Keywords: copper; electron transfer; heme; iron; magnesium; membrane-associated complex; metalloprotein; respiratory chain

F;63,377/Binding site: heme a iron (His) (axial ligands) #status predicted

F;239,289,290/Binding site: copper (His) #status predicted

F;239-243/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted

F;243/Binding site: oxygen (Tyr) #status predicted

F;375/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 52.4%; Score 43; DB 2; Length 790;

Best Local Similarity 72.7%; Pred. No. 29;  
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 PKLKVEVFPPF 15  
| |||| || |  
Db 650 PPLKVEYFPLP 660

RESULT 3

C84668

probable receptor-like protein kinase [imported] - *Arabidopsis thaliana*

C;Species: *Arabidopsis thaliana* (mouse-ear cress)

C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C;Accession: C84668

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.;  
Fujii, C.Y.; Mason, T.M.; Bowman, C.L.; Barnstead, M.E.; Feldblyum, T.V.; Buell,  
C.R.; Ketchum, K.A.; Lee, J.J.; Ronning, C.M.; Koo, H.; Moffat, K.S.; Cronin,  
L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.J.; Gill, J.E.; Adams,  
M.D.; Carrera, A.J.; Creasy, T.H.; Goodman, H.M.; Somerville, C.R.; Copenhaver,  
G.P.; Preuss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser,  
C.M.; Venter, J.C.

Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: C84668

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1007 <STO>

A;Cross-references: GB:AE002093; NID:g3885336; PIDN:AAC77864.1; GSPDB:GN00139

C;Genetics:

A;Gene: At2g27060

A;Map position: 2

Query Match 52.4%; Score 43; DB 2; Length 1007;  
Best Local Similarity 53.8%; Pred. No. 37;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 2 KPFPKLKVEVFPPF 14  
||:| || :|: |  
Db 898 KPYP SLKSDVYAF 910

RESULT 4

T45013

hypothetical protein [imported] - *Methanosarcina acetivorans* plasmid pC2A

C;Species: *Methanosarcina acetivorans*

C;Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000

C;Accession: T45013

R;Metcalf, W.W.; Zhang, J.K.; Apolinario, E.; Sowers, K.R.; Wolfe, R.S.  
Proc. Natl. Acad. Sci. U.S.A. 94, 2626-2631, 1997

A;Title: A genetic system for Archaea of the genus *Methanosarcina*: liposome-mediated transformation and construction of shuttle vectors.

A;Reference number: Z22897; MUID:97226004; PMID:9122246

A;Accession: T45013

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-190 <MET>  
A;Cross-references: EMBL:U78295; NID:g1763609; PIDN:AAB39747.1; PID:g1763613  
A;Experimental source: strain C2A  
C;Genetics:  
A;Genome: plasmid pC2A

Query Match 51.2%; Score 42; DB 2; Length 190;  
Best Local Similarity 64.3%; Pred. No. 9.7;  
Matches 9; Conservative 2; Mismatches 1; Indels 2; Gaps 1;

QY 4 FPKLKVEVFP--FP 15  
||||: |:|| ||  
Db 110 FPKLEKELFPEQFP 123

RESULT 5

T28908

hypothetical protein T26C11.2 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 18-Feb-2000

C;Accession: T28908

R;Martin, J.

submitted to the EMBL Data Library, December 1995

A;Description: The sequence of *C. elegans* cosmid T26C11.

A;Reference number: Z20542

A;Accession: T28908

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-343 <MAR>

A;Cross-references: EMBL:U41017; PIDN:AAC48211.1; GSPDB:GN00028; CESP:T26C11.2

A;Experimental source: strain Bristol N2; clone T26C11

C;Genetics:

A;Gene: CESP:T26C11.2

A;Map position: X

C;Superfamily: hydroxyproline-rich glycoprotein

Query Match 51.2%; Score 42; DB 2; Length 343;  
Best Local Similarity 64.3%; Pred. No. 18;  
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 KPFPKLVKVEVFPFP 15  
|| || | | || |  
Db 7 KPTPKPKSEFPKP 20

RESULT 6

AD2650

oxidoreductase ordL [imported] - *Agrobacterium tumefaciens* (strain C58, Dupont)

C;Species: *Agrobacterium tumefaciens*

C;Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002

C;Accession: AD2650

R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.; Kitajima, J.P.; Okura, V.K.; Almeida Jr., N.F.; Zhou, Y.; Bovee Sr., D.; Chapman, P.; Clendenning, J.; Deatherage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClelland, E.; Palmieri, A.; Raymond, C.; Rouse, G.; Saenphimmachak, C.; Wu, Z.; Gordon, D.; Eisen, J.A.; Paulsen, I.; Karp, P.; Romero, P.; Zhang, S.



Science 294, 2317-2323, 2001

A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, B.; Liao, L.; Kim, S.; Hendrick, C.; Zhao, Z.; Dolan, M.; Tingey, S.V.; Tomb, J.; Gordon, M.P.; Olson, M.V.; Nester, E.W.

A;Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.

A;Reference number: AB2577; MUID:21608550; PMID:11743193

A;Accession: AD2650

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-428 <KUR>

A;Cross-references: GB:AE008688; PIDN:AAL41618.1; PID:g17738956; GSPDB:GN00186

A;Experimental source: strain C58 (Dupont)

C;Genetics:

A;Gene: ordL

A;Map position: circular chromosome

C;Superfamily: hypothetical protein HI0499

Query Match 51.2%; Score 42; DB 2; Length 428;  
Best Local Similarity 61.5%; Pred. No. 23;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 LKPFPKLKVEVFP 13  
| || :||| ||  
Db 393 LAPFARLKVPAPF 405

#### RESULT 7

C97432

probable oxidoreductase ordL AGR\_C\_1066 [imported] - *Agrobacterium tumefaciens* (strain C58, Cereon)

C;Species: *Agrobacterium tumefaciens*

C;Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002

C;Accession: C97432

R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, B.S.; Cao, Y.; Askenazi, M.; Halling, C.; Mullin, L.; Houmiel, K.; Gordon, J.; Vaudin, M.; Iartchouk, O.; Epp, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Flanagan, C.; Crowell, C.; Gurson, J.; Lomo, C.; Sear, C.; Strub, G.; Cielo, C.; Slater, S.

Science 294, 2323-2328, 2001

A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tumefaciens* C58.

A;Reference number: A97359; MUID:21608551; PMID:11743194

A;Accession: C97432

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-428 <KUR>

A;Cross-references: GB:AE007869; PIDN:AAK86412.1; PID:g15155546; GSPDB:GN00169

C;Genetics:

A;Gene: AGR\_C\_1066

A;Map position: circular chromosome

C;Superfamily: hypothetical protein HI0499

Query Match 51.2%; Score 42; DB 2; Length 428;  
Best Local Similarity 61.5%; Pred. No. 23;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 LKPFPKLKVEVFP 13  
| || :||| ||  
Db 393 LAPFARLKVPAPFP 405

RESULT 8

T18974

hypothetical protein C06A1.4 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C;Accession: T18974

R;McMurray, A.

submitted to the EMBL Data Library, June 1995

A;Reference number: Z19054

A;Accession: T18974

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-939 <WIL>

A;Cross-references: EMBL:Z49886; PIDN:CAA90054.1; GSPDB:GN00020; CESP:C06A1.4

A;Experimental source: clone C06A1

C;Genetics:

A;Gene: CESP:C06A1.4

A;Map position: 2

A;Introns: 52/3; 116/2; 146/3; 282/1; 524/2; 583/1; 639/2; 697/3; 779/3; 901/2

Query Match 51.2%; Score 42; DB 2; Length 939;  
Best Local Similarity 61.5%; Pred. No. 51;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 LKPFPKLKVEVFP 13  
|| | :||| ||  
Db 162 LKSLPCIKLEVFP 174

RESULT 9

T22933

hypothetical protein F58G1.1 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C;Accession: T22933

R;Smye, R.

submitted to the EMBL Data Library, November 1996

A;Reference number: Z19639

A;Accession: T22933

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-965 <WIL>

A;Cross-references: EMBL:Z81556; PIDN:CAB04524.1; GSPDB:GN00020; CESP:F58G1.1

A;Experimental source: clone F58G1

C;Genetics:

A;Gene: CESP:F58G1.1

A;Map position: 2

A;Introns: 52/3; 116/2; 172/3; 308/1; 550/2; 609/1; 665/2; 723/3; 805/3; 927/2

Query Match 51.2%; Score 42; DB 2; Length 965;  
Best Local Similarity 61.5%; Pred. No. 53;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 LKPFPKLKVEVFP 13  
|| | :|:||||  
Db 188 LKSLPCIKLEVFP 200

RESULT 10

S44745

C02D5.3 protein - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 14-Sep-1994 #sequence\_revision 12-May-1995 #text\_change 09-Sep-1997

C;Accession: S44745

R;Du, Z.

submitted to the EMBL Data Library, May 1993

A;Description: Sequence of the C. elegans cosmid C02D5.

A;Reference number: S44613

A;Accession: S44745

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-379 <DUZ>

A;Cross-references: EMBL:L16622; NID:g289603; PID:g289606

C;Genetics:

A;Introns: 15/1; 90/1; 174/2; 196/1; 272/3

Query Match 50.6%; Score 41.5; DB 2; Length 379;  
Best Local Similarity 66.7%; Pred. No. 24;  
Matches 10; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

Qy 1 LKPFPKLKVEVFPPF 15  
:| || |||| || |  
Db 320 IKDFP-LKVESFPPG 333

RESULT 11

LRRTH

clathrin heavy chain - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 22-Jun-1999

C;Accession: A39941

R;Kirchhausen, T.; Harrison, S.C.; Ping Chow, E.; Mattaliano, R.J.;

Ramachandran, K.L.; Smart, J.; Brosius, J.

Proc. Natl. Acad. Sci. U.S.A. 84, 8805-8809, 1987

A;Title: Clathrin heavy chain: molecular cloning and complete primary structure.

A;Reference number: A39941; MUID:88097376; PMID:3480512

A;Accession: A39941

A;Molecule type: mRNA

A;Residues: 1-1675 <KIR>

A;Cross-references: GB:J03583; NID:g203301; PIDN:AAA40874.1; PID:g203302

C;Comment: Clathrin, the major protein component of coated pits and vesicles, is a three-legged, pinwheel-shaped structure. Each leg contains a heavy chain with a light chain noncovalently bonded near its carboxyl end. The heavy chains are also held together by noncovalent interactions.

C;Comment: The amino end of the mature protein is blocked.

C;Superfamily: clathrin heavy chain

C;Keywords: coated pits

F;1-479/Domain: amino-terminal <TER>

F;480-523/Region: link

F;524-634/Domain: distal <DIS>  
F;635-638/Region: joint #status predicted  
F;639-1675/Domain: proximal <PRX>

Query Match 50.6%; Score 41.5; DB 1; Length 1675;  
Best Local Similarity 64.3%; Pred. No. 1.1e+02;  
Matches 9; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

Qy 2 KPFPKCLKVEVFPPF 15  
:|||| |:|| ||  
Db 241 QPFPKKAVDVF-FP 253

RESULT 12

T38233

probable cystathionine gamma-synthase - fission yeast (*Schizosaccharomyces pombe*)

C;Species: *Schizosaccharomyces pombe*

C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jan-2000

C;Accession: T38233

R;Murphy, L.; Harris, D.; Wood, V.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, February 1998

A;Reference number: Z21780

A;Accession: T38233

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-398 <MUR>

A;Cross-references: EMBL:AL021813; PIDN:CAA16988.1; GSPDB:GN00066;

SPDB:SPAC23A1.14c

A;Experimental source: strain 972h-; cosmid c23A1

C;Genetics:

A;Gene: SPDB:SPAC23A1.14c

A;Map position: 1

C;Superfamily: O-succinylhomoserine (thiol)-lyase

Query Match 50.0%; Score 41; DB 2; Length 398;  
Best Local Similarity 57.1%; Pred. No. 31;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 LKPFPKCLKVEVFPPF 14  
|:|| || | ||:  
Db 52 LQPFTKLAEEEDFPY 65

RESULT 13

A82221

extracellular solute-binding protein, family 7 VC1273 [imported] - *Vibrio cholerae* (strain N16961 serogroup O1)

C;Species: *Vibrio cholerae*

C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

C;Accession: A82221

R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.;  
Dodson, R.J.; Haft, D.H.; Hickey, E.K.; Peterson, J.D.; Umayam, L.A.; Gill,  
S.R.; Nelson, K.E.; Read, T.D.; Tettelin, H.; Richardson, D.; Ermolaeva, M.D.;  
Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.; McDonald, L.;  
Utterback, T.; Fleishmann, R.D.; Nierman, W.C.; White, O.; Salzberg, S.L.;  
Smith, H.O.; Colwell, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A;Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.

A;Reference number: A82035; MUID:20406833; PMID:10952301

A;Accession: A82221

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-401 <HEI>

A;Cross-references: GB:AE004206; GB:AE003852; NID:g9655749; PIDN:AAF94432.1; GSPDB:GN00126; TIGR:VC1273

A;Experimental source: serogroup O1; strain N16961; biotype El Tor

C;Genetics:

A;Gene: VC1273

A;Map position: 1

Query Match 50.0%; Score 41; DB 2; Length 401;  
Best Local Similarity 58.3%; Pred. No. 31;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 4 FPKLKVEVFPFP 15  
|| :||: || |  
Db 336 FPDVKVKTFPAP 347

#### RESULT 14

S75790

hypothetical protein sll0827 - *Synechocystis* sp. (strain PCC 6803)

C;Species: *Synechocystis* sp.

A;Variety: PCC 6803

C;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000

C;Accession: S75790

R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugiura, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 3, 109-136, 1996

A;Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.

A;Reference number: S74322; MUID:97061201; PMID:8905231

A;Accession: S75790

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-496 <KAN>

A;Cross-references: EMBL:D64003; GB:AB001339; NID:g1001200; PIDN:BAA10525.1; PID:g1001279

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C;Superfamily: *Synechocystis* hypothetical protein sll0827

Query Match 50.0%; Score 41; DB 2; Length 496;  
Best Local Similarity 63.6%; Pred. No. 39;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 4 FPKLKVEVFPF 14  
||| | :|||  
Db 64 FPKRPVRIFPF 74

RESULT 15

C64432

hypothetical protein MJ1060 - Methanococcus jannaschii

C;Species: Methanococcus jannaschii

C;Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000

C;Accession: C64432

R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.A.; FitzGerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Scott, J.L.; Geoghagen, N.S.M.; Weidman, J.F.; Fuhrmann, J.L.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.

Science 273, 1058-1073, 1996

A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.

A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.

A;Reference number: A64300; MUID:96337999; PMID:8688087

A;Accession: C64432

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-537 <BUL>

A;Cross-references: GB:U67549; GB:L77117; NID:g2826363; PIDN:AAB99072.1;

PID:gl499906; TIGR:MJ1060

C;Genetics:

A;Map position: FOR1000459-1002072

Query Match 50.0%; Score 41; DB 2; Length 537;

Best Local Similarity 75.0%; Pred. No. 42;

Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 FPKLKVEVFPFP 15

||| || ||| |

Db 381 FPKDKVIVFPDP 392

Search completed: August 24, 2004, 15:52:59

Job time : 17.5522 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 24, 2004, 15:51:19 ; Search time 54.291 Seconds  
(without alignments)  
86.825 Million cell updates/sec

Title: US-09-641-801-8  
Perfect score: 82  
Sequence: 1 LKPFPKLKVEVFPPF 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1295152 seqs, 314255058 residues

Total number of hits satisfying chosen parameters: 1295152

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	
No.	Score Match Length DB ID	Description

1	82	100.0	15	14	US-10-281-652-8	Sequence 8, Appli
2	48.5	59.1	309	12	US-10-424-599-235668	Sequence 235668,
3	46	56.1	100	12	US-10-424-599-146821	Sequence 146821,
4	46	56.1	365	16	US-10-437-963-178880	Sequence 178880,
5	44.5	54.3	1173	16	US-10-437-963-172791	Sequence 172791,
6	43	52.4	954	16	US-10-771-931-3	Sequence 3, Appli
7	42.5	51.8	1515	16	US-10-437-963-137588	Sequence 137588,
8	42	51.2	85	12	US-10-424-599-266211	Sequence 266211,
9	42	51.2	232	15	US-10-104-047-3111	Sequence 3111, Ap
10	42	51.2	300	16	US-10-437-963-117046	Sequence 117046,
11	42	51.2	410	12	US-10-425-114-64363	Sequence 64363, A
12	42	51.2	443	12	US-10-425-114-45491	Sequence 45491, A
13	42	51.2	554	15	US-10-108-260A-2682	Sequence 2682, Ap
14	42	51.2	950	15	US-10-295-027-1340	Sequence 1340, Ap
15	42	51.2	1034	16	US-10-437-963-201195	Sequence 201195,
16	41	50.0	59	12	US-10-424-599-199084	Sequence 199084,
17	41	50.0	68	12	US-10-424-599-224398	Sequence 224398,
18	41	50.0	74	9	US-09-738-626-6030	Sequence 6030, Ap
19	41	50.0	90	12	US-10-424-599-274254	Sequence 274254,
20	41	50.0	188	10	US-09-791-932-90	Sequence 90, Appl
21	41	50.0	239	12	US-10-425-114-66814	Sequence 66814, A
22	41	50.0	333	12	US-10-424-599-213540	Sequence 213540,
23	41	50.0	374	16	US-10-437-963-118336	Sequence 118336,
24	41	50.0	764	16	US-10-437-963-170833	Sequence 170833,
25	41	50.0	842	16	US-10-437-963-193321	Sequence 193321,
26	40.5	49.4	1556	15	US-10-369-493-1796	Sequence 1796, Ap
27	40	48.8	105	16	US-10-437-963-144851	Sequence 144851,
28	40	48.8	106	12	US-10-424-599-150187	Sequence 150187,
29	40	48.8	121	12	US-10-424-599-218369	Sequence 218369,
30	40	48.8	131	16	US-10-437-963-138858	Sequence 138858,
31	40	48.8	171	12	US-10-425-114-48415	Sequence 48415, A
32	40	48.8	258	12	US-10-424-599-176441	Sequence 176441,
33	40	48.8	304	12	US-10-424-599-170626	Sequence 170626,
34	40	48.8	340	14	US-10-083-357-1337	Sequence 1337, Ap
35	40	48.8	496	12	US-10-425-114-40273	Sequence 40273, A
36	40	48.8	626	9	US-09-906-209-12	Sequence 12, Appl
37	40	48.8	626	9	US-09-906-209-13	Sequence 13, Appl
38	40	48.8	650	16	US-10-437-963-187658	Sequence 187658,
39	40	48.8	711	12	US-10-424-599-280508	Sequence 280508,
40	40	48.8	927	16	US-10-437-963-121736	Sequence 121736,
41	39	47.6	60	12	US-10-424-599-189179	Sequence 189179,
42	39	47.6	71	12	US-10-424-599-179022	Sequence 179022,
43	39	47.6	89	16	US-10-437-963-194732	Sequence 194732,
44	39	47.6	90	12	US-10-424-599-145940	Sequence 145940,
45	39	47.6	90	12	US-10-424-599-216387	Sequence 216387,

#### ALIGNMENTS

##### RESULT 1

US-10-281-652-8

; Sequence 8, Application US/10281652

; Publication No. US20030091606A1

; GENERAL INFORMATION:

; APPLICANT: STANTON, G. John



```

; APPLICANT: HUGHES, Thomas K.
; APPLICANT: BOLDOGH, Istvan
; TITLE OF INVENTION: USE OF COLOSTRININ, CONSTITUENT PEPTIDES THEREOF, AND
; TITLE OF INVENTION: ANALOGS THEREOF AS OXIDATIVE STRESS REGULATORS
; FILE REFERENCE: 265.00220101
; CURRENT APPLICATION NUMBER: US/10/281,652
; CURRENT FILING DATE: 2002-10-28
; PRIOR APPLICATION NUMBER: US/09/641,803
; PRIOR FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: 60/149,310
; PRIOR FILING DATE: 1999-08-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: peptide
US-10-281-652-8

```

```

Query Match          100.0%; Score 82; DB 14; Length 15;
Best Local Similarity 100.0%; Pred. No. 2.9e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 LKPFPKLKVEVFPPF 15
        |||||
Db      1 LKPFPKLKVEVFPPF 15

```

## RESULT 2

US-10-424-599-235668

```

; Sequence 235668, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated
With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 235668
; LENGTH: 309
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(309)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_54835C.1.pep

```

US-10-424-599-235668

Query Match 59.1%; Score 48.5; DB 12; Length 309;  
Best Local Similarity 50.0%; Pred. No. 13;  
Matches 12; Conservative 2; Mismatches 1; Indels 9; Gaps 2;

Qy 1 LKPF-PKLVKEVF-----PFP 15  
|||| |: :||| |||  
Db 275 LKPFAPKIPFIEVFLEAIKPTLPFP 298

RESULT 3

US-10-424-599-146821  
; Sequence 146821, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated  
With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 146821  
; LENGTH: 100  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_103599C.1.pep  
US-10-424-599-146821

Query Match 56.1%; Score 46; DB 12; Length 100;  
Best Local Similarity 72.7%; Pred. No. 10;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 PFPKLVKEVFP 13  
||||:|:| |  
Db 79 PFPKIKVKVSP 89

RESULT 4

US-10-437-963-178880  
; Sequence 178880, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping

```
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules
Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 178880
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_76396C.1.pep
US-10-437-963-178880
```

```
Query Match          56.1%; Score 46; DB 16; Length 365;
Best Local Similarity 53.3%; Pred. No. 38;
Matches      8; Conservative      3; Mismatches      4; Indels      0; Gaps      0;
```

```
Qy      1 LKPFPKLKVEVFPFP 15
        | ||||: |:|: |
Db      19 LLPFPKVSQVYTVP 33
```

RESULT 5

```
US-10-437-963-172791
; Sequence 172791, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules
Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 172791
; LENGTH: 1173
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1173)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_70893C.1.pep
US-10-437-963-172791
```

```
Query Match          54.3%; Score 44.5; DB 16; Length 1173;
```

Best Local Similarity 60.0%; Pred. No. 2.2e+02;  
Matches 9; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

Qy 2 KPFPKLKVEVF-PFP 15  
:||||| : | |||  
Db 968 RFPFKLAFKYFGPFP 982

RESULT 6

US-10-771-931-3  
; Sequence 3, Application US/10771931  
; Publication No. US20040151737A1  
; GENERAL INFORMATION:  
; APPLICANT: Courtney, Harry  
; TITLE OF INVENTION: Streptococcal Serum Opacity Factors And Fibronectin-Binding Proteins And  
; TITLE OF INVENTION: Peptides Thereof For The Treatment And Detection of Streptococcal Infection  
; FILE REFERENCE: 13314.1001U  
; CURRENT APPLICATION NUMBER: US/10/771,931  
; CURRENT FILING DATE: 2004-02-04  
; NUMBER OF SEQ ID NOS: 57  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 954  
; TYPE: PRT  
; ORGANISM: Streptococcus pyogenes  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(954)  
; OTHER INFORMATION: Xaa = Any Amino Acid  
US-10-771-931-3

Query Match 52.4%; Score 43; DB 16; Length 954;  
Best Local Similarity 53.8%; Pred. No. 3.1e+02;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 3 PFPKLKVEVFPFP 15  
| | : | : | | |  
Db 730 PIPELDIEVVPIF 742

RESULT 7

US-10-437-963-137588  
; Sequence 137588, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With

```

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 137588
; LENGTH: 1515
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1515)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_39057C.1.pep
US-10-437-963-137588

```

```

Query Match          51.8%; Score 42.5; DB 16; Length 1515;
Best Local Similarity 60.0%; Pred. No. 6e+02;
Matches      9; Conservative      2; Mismatches      3; Indels      1; Gaps      1;

```

```

QY          2 KPFPKLVKVEVF-PFP 15
              :||||| : | |||
Db          1374 QPFPKLVFKYFGPFP 1388

```

```

RESULT 8
US-10-424-599-266211
; Sequence 266211, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated
With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 266211
; LENGTH: 85
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_8240C.1.pep
US-10-424-599-266211

```

```

Query Match          51.2%; Score 42; DB 12; Length 85;
Best Local Similarity 58.3%; Pred. No. 36;
Matches      7; Conservative      3; Mismatches      2; Indels      0; Gaps      0;

```

```

QY          1 LKFPKLVKVEVF 12
              : || ||: ||: |
Db          13 ISPFKLVQVEIF 24

```

RESULT 9

US-10-104-047-3111  
 ; Sequence 3111, Application US/10104047  
 ; Publication No. US20030236392A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HELIX RESEARCH INSTITUTE  
 ; TITLE OF INVENTION: No. US20030236392A1e1 full length cDNA  
 ; FILE REFERENCE: H1-A0105  
 ; CURRENT APPLICATION NUMBER: US/10/104,047  
 ; CURRENT FILING DATE: 2002-03-25  
 ; PRIOR APPLICATION NUMBER:  
 ; PRIOR FILING DATE:  
 ; NUMBER OF SEQ ID NOS: 4096  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 3111  
 ; LENGTH: 232  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-104-047-3111

Query Match 51.2%; Score 42; DB 15; Length 232;  
 Best Local Similarity 53.3%; Pred. No. 1e+02;  
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 LKFPKCLKVEVFPFP 15  
 | : | | | : | | |  
 Db 207 LRKFPVLPVHPWFPF 221

RESULT 10

US-10-437-963-117046  
 ; Sequence 117046, Application US/10437963  
 ; Publication No. US20040123343A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa, Thomas J.  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Wu, Wei  
 ; APPLICANT: Boukharov, Andrey A.  
 ; APPLICANT: Barbazuk, Brad  
 ; APPLICANT: Li, Ping  
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules  
 Associated With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53221)B  
 ; CURRENT APPLICATION NUMBER: US/10/437,963  
 ; CURRENT FILING DATE: 2003-05-14  
 ; NUMBER OF SEQ ID NOS: 204966  
 ; SEQ ID NO 117046  
 ; LENGTH: 300  
 ; TYPE: PRT  
 ; ORGANISM: Oryza sativa  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_2048C.1.pep

US-10-437-963-117046

Query Match 51.2%; Score 42; DB 16; Length 300;  
Best Local Similarity 54.5%; Pred. No. 1.3e+02;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 4 FPKLKVEVFPP 14  
|| |:|::||  
Db 46 FPSLRFEIYPF 56

RESULT 11

US-10-425-114-64363  
; Sequence 64363, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated  
With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 64363  
; LENGTH: 410  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3689-218-D2\_FLI.pep  
US-10-425-114-64363

Query Match 51.2%; Score 42; DB 12; Length 410;  
Best Local Similarity 54.5%; Pred. No. 1.9e+02;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 4 FPKLKVEVFPP 14  
|| |:|::||  
Db 159 FPSLRFEIYPF 169

RESULT 12

US-10-425-114-45491  
; Sequence 45491, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei

```
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated
With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 45491
; LENGTH: 443
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700377865_FLI.pep
US-10-425-114-45491
```

```
Query Match          51.2%; Score 42; DB 12; Length 443;
Best Local Similarity 54.5%; Pred. No. 2e+02;
Matches      6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy      4 FPKLKVEVFPF 14
        || |:|::||
Db      193 FPSLRFEIYPF 203
```

#### RESULT 13

```
US-10-108-260A-2682
; Sequence 2682, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1e1 full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2682
; LENGTH: 554
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-2682
```

```
Query Match          51.2%; Score 42; DB 15; Length 554;
Best Local Similarity 66.7%; Pred. No. 2.5e+02;
Matches      8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy      1 LKPFPPKLKVEVF 12
        | | | | | | | |
Db      83 LDPLPSLKVEVY 94
```

#### RESULT 14

```
US-10-295-027-1340
; Sequence 1340, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
```



```

; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynne, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1340
;   LENGTH: 950
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-295-027-1340

```

```

Query Match          51.2%;   Score 42;   DB 15;   Length 950;
Best Local Similarity 66.7%;   Pred. No. 4.4e+02;
Matches      8;   Conservative    1;   Mismatches    3;   Indels      0;   Gaps      0;

Qy          1 LKPFPKLKVEVF 12
             | | | | | | |
Db          479 LDPLPSLKVEVY 490

```

```

RESULT 15
US-10-437-963-201195
; Sequence 201195, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:

```

```

; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules
Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 201195
; LENGTH: 1034
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_96595C.1.pep
US-10-437-963-201195

```

```

Query Match          51.2%; Score 42; DB 16; Length 1034;
Best Local Similarity 66.7%; Pred. No. 4.9e+02;
Matches      8; Conservative      1; Mismatches      3; Indels      0; Gaps      0;

```

```

Qy      4 FPKLKVEVFPPF 15
        ||:| | || ||
Db      28 FPRLSVAVFYFP 39

```

```

Search completed: August 24, 2004, 16:41:26
Job time : 55.291 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 24, 2004, 15:23:00 ; Search time 46.3433 Seconds  
(without alignments)  
102.124 Million cell updates/sec

Title: US-09-641-801-8  
Perfect score: 82  
Sequence: 1 LKPFPKLKVEVFPFP 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	
No.	Score	Match Length DB ID Description
-----		

1	47	57.3	435	5	Q9VVQ7	Q9vvq7 drosophila
2	47	57.3	2443	4	Q96JI7	Q96ji7 homo sapien
3	46	56.1	159	10	Q948M2	Q948m2 oryza merid
4	46	56.1	161	10	Q93VC5	Q93vc5 oryza sativ
5	46	56.1	161	10	Q948M3	Q948m3 oryza glabe
6	46	56.1	161	10	Q93VC4	Q93vc4 oryza rufip
7	46	56.1	161	10	Q948M0	Q948m0 oryza glumi
8	46	56.1	161	10	Q948M1	Q948m1 oryza barth
9	46	56.1	163	10	Q948M4	Q948m4 oryza rufip
10	46	56.1	164	10	Q948L6	Q948l6 oryza rufip
11	46	56.1	164	10	Q948L7	Q948l7 oryza rufip
12	46	56.1	164	10	Q93W90	Q93w90 oryza sativ
13	46	56.1	164	10	Q948L9	Q948l9 oryza sativ
14	46	56.1	164	10	Q948L8	Q948l8 oryza rufip
15	46	56.1	365	10	Q948L5	Q948l5 oryza sativ
16	46	56.1	365	10	Q948L4	Q948l4 oryza sativ
17	46	56.1	367	10	Q9XIV7	Q9xiv7 oryza sativ
18	44	53.7	650	3	Q8J2K1	Q8j2k1 pichia angu
19	43	52.4	101	11	Q9CYU9	Q9cyu9 mus musculus
20	43	52.4	127	10	O22688	O22688 arabidopsis
21	43	52.4	213	16	Q8DT85	Q8dt85 streptococc
22	43	52.4	249	10	Q9LHC4	Q9lhc4 arabidopsis
23	43	52.4	790	17	Q97UN0	Q97un0 sulfolobus
24	43	52.4	1007	10	Q9ZVD4	Q9zvd4 arabidopsis
25	42.5	51.8	148	10	Q7XSE5	Q7xse5 oryza sativ
26	42	51.2	190	1	P94913	P94913 methanosarc
27	42	51.2	232	4	Q8NAJ2	Q8naj2 homo sapien
28	42	51.2	289	16	Q99ZW1	Q99zw1 streptococc
29	42	51.2	343	5	Q22807	Q22807 caenorhabdi
30	42	51.2	428	16	Q8UHS8	Q8uhs8 agrobacteri
31	42	51.2	554	4	Q8N1Y2	Q8nly2 homo sapien
32	42	51.2	762	12	Q993B3	Q993b3 simian cyto
33	42	51.2	939	5	Q17685	Q17685 caenorhabdi
34	42	51.2	965	5	O62275	O62275 caenorhabdi
35	42	51.2	1034	10	Q7XW39	Q7xw39 oryza sativ
36	41.5	50.6	1639	4	Q86TF2	Q86tf2 homo sapien
37	41.5	50.6	1675	13	Q8UUR1	Q8uur1 gallus gall
38	41.5	50.6	1684	11	Q80U89	Q80u89 mus musculus
39	41	50.0	70	16	Q8G2Z3	Q8g2z3 brucella su
40	41	50.0	182	16	Q8ZNB1	Q8znb1 salmonella
41	41	50.0	401	16	Q9KSI3	Q9ksi3 vibrio chol
42	41	50.0	466	10	Q9LSM0	Q9lsm0 arabidopsis
43	41	50.0	466	10	Q8GX09	Q8gx09 arabidopsis
44	41	50.0	477	3	Q9C2H5	Q9c2h5 neurospora
45	41	50.0	496	16	Q55425	Q55425 synechocyst

#### ALIGNMENTS

##### RESULT 1

Q9VVQ7

ID Q9VVQ7 PRELIMINARY; PRT; 435 AA.  
AC Q9VVQ7;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE CG18231 protein.  
 GN CG18231.  
 OS *Drosophila melanogaster* (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; *Drosophila*.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
 RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,

RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanenvong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of Drosophila melanogaster genome."  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of Drosophila melanogaster genome."  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AE003520; AAF49253.2; -.  
 DR FlyBase; FBgn0036796; CG18231.  
 SQ SEQUENCE 435 AA; 50674 MW; 0FB957A50F29AD34 CRC64;

Query Match 57.3%; Score 47; DB 5; Length 435;  
 Best Local Similarity 80.0%; Pred. No. 10;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 4 FPKLKVEVFP 13  
 |||||: |||  
 Db 395 FPKLKISVFP 404

## RESULT 2

Q96JI7

ID Q96JI7 PRELIMINARY; PRT; 2443 AA.  
 AC Q96JI7;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein KIAA1840 (Fragment).  
 GN KIAA1840.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;

RX MEDLINE=21245130; PubMed=11347906;  
 RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XX.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large Proteins in vitro.";  
 RL DNA Res. 8:85-95(2001).  
 DR EMBL; AB058743; BAB47469.2; -.  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003824; F:catalytic activity; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR GO; GO:0008199; F:ferric iron binding; IEA.  
 DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.  
 DR GO; GO:0006725; P:aromatic compound metabolism; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro; IPR000627; Dioxxygenase.  
 DR InterPro; IPR001360; Glyco\_hydro\_1.  
 DR InterPro; IPR001005; Myb\_DNA\_binding.  
 DR PROSITE; PS00572; GLYCOSYL\_HYDROL\_F1\_1; 1.  
 DR PROSITE; PS00037; MYB\_1; 1.  
 KW Hypothetical protein.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 2443 AA; 278805 MW; 580B24253D940D1E CRC64;

Query Match 57.3%; Score 47; DB 4; Length 2443;  
 Best Local Similarity 61.5%; Pred. No. 57;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 LKPFPKLKVEVFP 13  
 | | : | | | : | |  
 Db 1120 LTPYPKLKTALFP 1132

# RESULT 3

Q948M2

ID Q948M2 PRELIMINARY; PRT; 159 AA.  
 AC Q948M2;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Transcription factor OSH3 (Fragment).  
 GN OSH3.  
 OS Oryza meridionalis.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzeae; Oryza.  
 OX NCBI\_TaxID=40149;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. W1629;  
 RA Sato Y., Fukuda Y., Hirano H.;  
 RT "Evidence for the selection at KNOTTED1-like homobox gene, OSH3, locus  
 RT in rice.";  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AB071627; BAB68273.1; -.  
 DR Gramene; Q948M2; -.  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.

DR InterPro; IPR005540; KNOX1.  
DR InterPro; IPR005541; KNOX2.  
DR Pfam; PF03790; KNOX1; 1.  
DR Pfam; PF03791; KNOX2; 1.  
FT NON\_TER 1 1  
FT NON\_TER 159 159  
SQ SEQUENCE 159 AA; 16897 MW; 48753DB82F95AC16 CRC64;

Query Match 56.1%; Score 46; DB 10; Length 159;  
Best Local Similarity 53.3%; Pred. No. 5.8;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 LKPFPKLKVEVFPFP 15  
| ||||: |:|: |  
Db 19 LLPFPKVSQVYTVP 33

#### RESULT 4

Q93VC5

ID Q93VC5 PRELIMINARY; PRT; 161 AA.  
AC Q93VC5;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Transcription factor OSH3 (Fragment).  
GN OSH3.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzeae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. 451, and cv. Kasalath;  
RA Sato Y., Fukuda Y., Hirano H.;  
RT "Evidence for the selection at KNOTTED1-like homobox gene, OSH3, locus  
RT in rice.";  
RL Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AB071642; BAB68288.1; -.  
DR EMBL; AB071653; BAB68299.1; -.  
DR Gramene; Q93VC5; -.  
DR GO; GO:0005634; C:nucleus; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR InterPro; IPR005540; KNOX1.  
DR InterPro; IPR005541; KNOX2.  
DR Pfam; PF03790; KNOX1; 1.  
DR Pfam; PF03791; KNOX2; 1.  
FT NON\_TER 1 1  
FT NON\_TER 161 161  
SQ SEQUENCE 161 AA; 17174 MW; D0C0D61C8369FA77 CRC64;

Query Match 56.1%; Score 46; DB 10; Length 161;  
Best Local Similarity 53.3%; Pred. No. 5.9;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 LKPFPKLKVEVFPFP 15  
| ||||: |:|: |



Db 19 LLPFPKVSQVYTVP 33

RESULT 5

Q948M3

ID Q948M3 PRELIMINARY; PRT; 161 AA.  
AC Q948M3;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Transcription factor OSH3 (Fragment).  
GN OSH3.  
OS Oryza glaberrima (African rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzeae; Oryza.  
OX NCBI\_TaxID=4538;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. W440;  
RA Sato Y., Fukuda Y., Hirano H.;  
RT "Evidence for the selection at KNOTTED1-like homobox gene, OSH3, locus  
RT in rice."  
RL Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AB071626; BAB68272.1; -.  
DR Gramene; Q948M3; -.  
DR GO; GO:0005634; C:nucleus; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR InterPro; IPR005540; KNOX1.  
DR InterPro; IPR005541; KNOX2.  
DR Pfam; PF03790; KNOX1; 1.  
DR Pfam; PF03791; KNOX2; 1.  
FT NON\_TER 1 1  
FT NON\_TER 161 161  
SQ SEQUENCE 161 AA; 17116 MW; ECEA5658836DBA30 CRC64;

Query Match 56.1%; Score 46; DB 10; Length 161;  
Best Local Similarity 53.3%; Pred. No. 5.9;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 LKPFPLKVEVFPFP 15  
| ||||: |:|: |  
Db 19 LLPFPKVSQVYTVP 33

RESULT 6

Q93VC4

ID Q93VC4 PRELIMINARY; PRT; 161 AA.  
AC Q93VC4;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Transcription factor OSH3 (Fragment).  
GN OSH3.  
OS Oryza rufipogon (Wild rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC Ehrhartoideae; Oryzeae; Oryza.  
 OX NCBI\_TaxID=4529;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. W1680, cv. W1865, and cv. W1987;  
 RA Sato Y., Fukuda Y., Hirano H.;  
 RT "Evidence for the selection at KNOTTED1-like homobox gene, OSH3, locus  
 RT in rice.";  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AB071658; BAB68304.1; -.  
 DR EMBL; AB071660; BAB68306.1; -.  
 DR EMBL; AB071661; BAB68307.1; -.  
 DR Gramene; Q93VC4; -.  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR InterPro; IPR005540; KNOX1.  
 DR InterPro; IPR005541; KNOX2.  
 DR Pfam; PF03790; KNOX1; 1.  
 DR Pfam; PF03791; KNOX2; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 161 161  
 SQ SEQUENCE 161 AA; 17174 MW; D0C0D61C8369FA77 CRC64;

Query Match 56.1%; Score 46; DB 10; Length 161;  
 Best Local Similarity 53.3%; Pred. No. 5.9;  
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 LKPFPKLKVEVFPPF 15  
 | ||||: |:|: |  
 Db 19 LLPFPKVSQVYTVP 33

#### RESULT 7

Q948M0

ID Q948M0 PRELIMINARY; PRT; 161 AA.  
 AC Q948M0;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Transcription factor OSH3 (Fragment).  
 GN OSH3.  
 OS Oryza glumipatula.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzeae; Oryza.  
 OX NCBI\_TaxID=40148;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. W1187;  
 RA Sato Y., Fukuda Y., Hirano H.;  
 RT "Evidence for the selection at KNOTTED1-like homobox gene, OSH3, locus  
 RT in rice.";  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AB071629; BAB68275.1; -.  
 DR Gramene; Q948M0; -.  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.

DR InterPro; IPR005540; KNOX1.  
 DR InterPro; IPR005541; KNOX2.  
 DR Pfam; PF03790; KNOX1; 1.  
 DR Pfam; PF03791; KNOX2; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 161 161  
 SQ SEQUENCE 161 AA; 17132 MW; ECF23652E37FBA30 CRC64;

Query Match 56.1%; Score 46; DB 10; Length 161;  
 Best Local Similarity 53.3%; Pred. No. 5.9;  
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 LKPFPKLKVEVFPFP 15  
 | ||||: |:|: |  
 Db 19 LLPFPKVSQVYTVP 33

# RESULT 8

Q948M1

ID Q948M1 PRELIMINARY; PRT; 161 AA.  
 AC Q948M1;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Transcription factor OSH3 (Fragment).  
 GN OSH3.  
 OS Oryza barthii.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartioideae; Oryzeae; Oryza.  
 OX NCBI\_TaxID=65489;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. W1468;  
 RA Sato Y., Fukuda Y., Hirano H.;  
 RT "Evidence for the selection at KNOTTED1-like homobox gene, OSH3, locus  
 RT in rice."  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AB071628; BAB68274.1; -.  
 DR Gramene; Q948M1; -.  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR InterPro; IPR005540; KNOX1.  
 DR InterPro; IPR005541; KNOX2.  
 DR Pfam; PF03790; KNOX1; 1.  
 DR Pfam; PF03791; KNOX2; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 161 161  
 SQ SEQUENCE 161 AA; 17116 MW; ECEA5658836DBA30 CRC64;

Query Match 56.1%; Score 46; DB 10; Length 161;  
 Best Local Similarity 53.3%; Pred. No. 5.9;  
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 LKPFPKLKVEVFPFP 15  
 | ||||: |:|: |  
 Db 19 LLPFPKVSQVYTVP 33

RESULT 9

Q948M4

ID Q948M4 PRELIMINARY; PRT; 163 AA.  
 AC Q948M4;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Transcription factor OSH3 (Fragment).  
 GN OSH3.  
 OS Oryza rufipogon (Wild rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzeae; Oryza.  
 OX NCBI\_TaxID=4529;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. W629;  
 RA Sato Y., Fukuda Y., Hirano H.;  
 RT "Evidence for the selection at KNOTTED1-like homobox gene, OSH3, locus  
 RT in rice."  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AB071625; BAB68271.1; -.  
 DR Gramene; Q948M4; -.  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR InterPro; IPR005540; KNOX1.  
 DR InterPro; IPR005541; KNOX2.  
 DR Pfam; PF03790; KNOX1; 1.  
 DR Pfam; PF03791; KNOX2; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 163 163  
 SQ SEQUENCE 163 AA; 17248 MW; 3E4D33040B16B866 CRC64;

Query Match 56.1%; Score 46; DB 10; Length 163;  
 Best Local Similarity 53.3%; Pred. No. 5.9;  
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 LKPFPPKLKVEVFPPF 15  
 | ||||: |:|: |  
 Db 18 LLPFPKVSQVYTVP 32

RESULT 10

Q948L6

ID Q948L6 PRELIMINARY; PRT; 164 AA.  
 AC Q948L6;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Transcription factor OSH3 (Fragment).  
 GN OSH3.  
 OS Oryza rufipogon (Wild rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzeae; Oryza.

OX NCBI\_TaxID=4529;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. W630;  
 RA Sato Y., Fukuda Y., Hirano H.;  
 RT "Evidence for the selection at KNOTTED1-like homobox gene, OSH3, locus  
 RT in rice."  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AB071662; BAB68308.1; -.  
 DR Gramene; Q948L6; -.  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR InterPro; IPR005540; KNOX1.  
 DR InterPro; IPR005541; KNOX2.  
 DR Pfam; PF03790; KNOX1; 1.  
 DR Pfam; PF03791; KNOX2; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 164 164  
 SQ SEQUENCE 164 AA; 17432 MW; 945B9D2936C8D45C CRC64;

Query Match 56.1%; Score 46; DB 10; Length 164;  
 Best Local Similarity 53.3%; Pred. No. 6;  
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 LKPFPKLKVEVFPFP 15  
 | ||||: |:|: |  
 Db 19 LLPFPKVSQVYTVP 33

# RESULT 11

Q948L7

ID Q948L7 PRELIMINARY; PRT; 164 AA.  
 AC Q948L7;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created).  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Transcription factor OSH3 (Fragment).  
 GN OSH3.  
 OS Oryza rufipogon (Wild rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzeae; Oryza.  
 OX NCBI\_TaxID=4529;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. W1811;  
 RA Sato Y., Fukuda Y., Hirano H.;  
 RT "Evidence for the selection at KNOTTED1-like homobox gene, OSH3, locus  
 RT in rice."  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AB071659; BAB68305.1; -.  
 DR Gramene; Q948L7; -.  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR InterPro; IPR005540; KNOX1.  
 DR InterPro; IPR005541; KNOX2.  
 DR Pfam; PF03790; KNOX1; 1.

DR Pfam; PF03791; KNOX2; 1.  
FT NON\_TER 1 1  
FT NON\_TER 164 164  
SQ SEQUENCE 164 AA; 17433 MW; 430B9D261998B6C5 CRC64;

Query Match 56.1%; Score 46; DB 10; Length 164;  
Best Local Similarity 53.3%; Pred. No. 6;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 LKPFPKLKVEVFPPF 15  
| ||||: |:|: |  
Db 19 LLPFPKVSQVYTVP 33

RESULT 12

Q93W90

ID Q93W90 PRELIMINARY; PRT; 164 AA.  
AC Q93W90;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Transcription factor OSH3 (Fragment).  
GN OSH3.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzeae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Various strains;  
RA Sato Y., Fukuda Y., Hirano H.;  
RT "Evidence for the selection at KNOTTED1-like homobox gene, OSH3, locus  
in rice."  
RL Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AB071630; BAB68276.1; -.  
DR EMBL; AB071631; BAB68277.1; -.  
DR EMBL; AB071632; BAB68278.1; -.  
DR EMBL; AB071633; BAB68279.1; -.  
DR EMBL; AB071634; BAB68280.1; -.  
DR EMBL; AB071635; BAB68281.1; -.  
DR EMBL; AB071636; BAB68282.1; -.  
DR EMBL; AB071637; BAB68283.1; -.  
DR EMBL; AB071638; BAB68284.1; -.  
DR EMBL; AB071639; BAB68285.1; -.  
DR EMBL; AB071640; BAB68286.1; -.  
DR EMBL; AB071641; BAB68287.1; -.  
DR EMBL; AB071643; BAB68289.1; -.  
DR EMBL; AB071644; BAB68290.1; -.  
DR EMBL; AB071645; BAB68291.1; -.  
DR EMBL; AB071646; BAB68292.1; -.  
DR EMBL; AB071647; BAB68293.1; -.  
DR EMBL; AB071648; BAB68294.1; -.  
DR EMBL; AB071650; BAB68296.1; -.  
DR EMBL; AB071651; BAB68297.1; -.  
DR EMBL; AB071652; BAB68298.1; -.  
DR EMBL; AB071654; BAB68300.1; -.

DR EMBL; AB071655; BAB68301.1; -.  
 DR EMBL; AB071656; BAB68302.1; -.  
 DR Gramene; Q93W90; -.  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR InterPro; IPR005540; KNOX1.  
 DR InterPro; IPR005541; KNOX2.  
 DR Pfam; PF03790; KNOX1; 1.  
 DR Pfam; PF03791; KNOX2; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 164 164  
 SQ SEQUENCE 164 AA; 17403 MW; 430B9D3BAF43C6C5 CRC64;

Query Match 56.1%; Score 46; DB 10; Length 164;  
 Best Local Similarity 53.3%; Pred. No. 6;  
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 LKPFPKLVKVEVFPFP 15  
 | ||||: |:|: |  
 Db 19 LLPFPKVSQVYTV 33

# RESULT 13

Q948L9

ID Q948L9 PRELIMINARY; PRT; 164 AA.  
 AC Q948L9;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Transcription factor OSH3 (Fragment).  
 GN OSH3.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzeae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. 868;  
 RA Sato Y., Fukuda Y., Hirano H.;  
 RT "Evidence for the selection at KNOTTED1-like homobox gene, OSH3, locus  
 RT in rice."  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AB071649; BAB68295.1; -.  
 DR Gramene; Q948L9; -.  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR InterPro; IPR005540; KNOX1.  
 DR InterPro; IPR005541; KNOX2.  
 DR Pfam; PF03790; KNOX1; 1.  
 DR Pfam; PF03791; KNOX2; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 164 164  
 SQ SEQUENCE 164 AA; 17433 MW; 4317F14BAF43D3D0 CRC64;

Query Match 56.1%; Score 46; DB 10; Length 164;  
 Best Local Similarity 53.3%; Pred. No. 6;

Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```
Qy      1 LKPFPKLKVEVFPPF 15
          | ||||: |:|: |
Db      19 LLPFPKVSQVYTVP 33
```

RESULT 14

Q948L8

ID Q948L8 PRELIMINARY; PRT; 164 AA.  
AC Q948L8;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Transcription factor OSH3 (Fragment).  
GN OSH3.  
OS Oryza rufipogon (Wild rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzeae; Oryza.  
OX NCBI\_TaxID=4529;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. W137;  
RA Sato Y., Fukuda Y., Hirano H.;  
RT "Evidence for the selection at KNOTTED1-like homobox gene, OSH3, locus  
RT in rice.";  
RL Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AB071657; BAB68303.1; -.  
DR Gramene; Q948L8; -.  
DR GO; GO:0005634; C:nucleus; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR InterPro; IPR005540; KNOX1.  
DR InterPro; IPR005541; KNOX2.  
DR Pfam; PF03790; KNOX1; 1.  
DR Pfam; PF03791; KNOX2; 1.  
FT NON\_TER 1 1  
FT NON\_TER 164 164  
SQ SEQUENCE 164 AA; 17403 MW; 430B9D3BAF43C6C5 CRC64;

Query Match 56.1%; Score 46; DB 10; Length 164;  
Best Local Similarity 53.3%; Pred. No. 6;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```
Qy      1 LKPFPKLKVEVFPPF 15
          | ||||: |:|: |
Db      19 LLPFPKVSQVYTVP 33
```

RESULT 15

Q948L5

ID Q948L5 PRELIMINARY; PRT; 365 AA.  
AC Q948L5;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Transcription factor OSH3.



GN OSH3.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzeae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Sato Y., Fukuda Y., Hirano H.;  
 RT "Evidence for the selection at KNOTTED1-like homobox gene, OSH3, locus  
 RT in rice."  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AB071663; BAB68309.1; -.  
 DR Gramene; Q948L5; -.  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003700; F:transcription factor activity; IEA.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
 DR InterPro; IPR005539; ELK.  
 DR InterPro; IPR001356; Homeobox.  
 DR InterPro; IPR005540; KNOX1.  
 DR InterPro; IPR005541; KNOX2.  
 DR Pfam; PF03789; ELK; 1.  
 DR Pfam; PF03790; KNOX1; 1.  
 DR Pfam; PF03791; KNOX2; 1.  
 DR ProDom; PD000010; Homeobox; 1.  
 DR SMART; SM00389; HOX; 1.  
 DR PROSITE; PS50071; HOMEBOX\_2; 1.  
 SQ SEQUENCE 365 AA; 40037 MW; FFC5346C7B5210DE CRC64;

Query Match 56.1%; Score 46; DB 10; Length 365;  
 Best Local Similarity 53.3%; Pred. No. 13;  
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 LKPFPKLKVEVFPPF 15  
 | ||||: |:|: |  
 Db 19 LLPFPKVSQVYTV 33

Search completed: August 24, 2004, 15:50:55  
 Job time : 50.3433 secs

OM protein - protein search, using sw model

Run on: August 24, 2004, 14:57:04 ; Search time 8.0597 Seconds  
(without alignments)  
96.908 Million cell updates/sec

Title: US-09-641-801-8  
Perfect score: 82  
Sequence: 1 LKPFPLKVEVFPFP 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	%		DB	ID	Description
		Query Match	Length			
1	41.5	50.6	1675	1	CLH1_HUMAN	Q00610 homo sapien
2	41.5	50.6	1675	1	CLH_BOVIN	P49951 bos taurus
3	41.5	50.6	1675	1	CLH_RAT	P11442 rattus norv
4	41	50.0	398	1	YFHE_SCHPO	O42851 schizosacch
5	41	50.0	537	1	YA60_METJA	Q58460 methanococc
6	40.5	49.4	835	1	TLR4_RAT	Q9qx05 rattus norv
7	40	48.8	280	1	COBS_METMA	Q8pzb4 methanosarc
8	40	48.8	304	1	TYSY_YEAST	P06785 saccharomyc
9	40	48.8	626	1	PGMP_PEA	Q9sm59 pisum sativ
10	39	47.6	307	1	TYSY_MOUSE	P07607 mus musculu
11	39	47.6	327	1	GDB2_WHEAT	P08453 triticum ae
12	39	47.6	330	1	CD22_PONPY	Q9nle3 pongo pygma
13	39	47.6	479	1	YEBU_ECOLI	P76273 escherichia
14	39	47.6	681	1	SSAV_SALTY	P74856 salmonella
15	39	47.6	782	1	Y044_UREPA	Q9pral ureaplasma
16	38.5	47.0	258	1	FLIR_BUCAP	Q8ka35 buchnera ap
17	38.5	47.0	583	1	KPYA_RICCO	Q43117 ricinus com

18	38	46.3	30	1	YVEF_BACSU	007004	bacillus su
19	38	46.3	98	1	YBGA_THEMA	Q08640	thermotoga
20	38	46.3	131	1	WFD6_HUMAN	Q9bqy6	homo sapien
21	38	46.3	542	1	ZYX_CHICK	Q04584	gallus gall
22	38	46.3	570	1	PUR6_CANGA	O74197	candida gla
23	38	46.3	607	1	VATA_NEUCR	P11592	neurospora
24	38	46.3	621	1	ACD9_HUMAN	Q9h845	homo sapien
25	38	46.3	632	1	Z179_HUMAN	Q9ulx5	homo sapien
26	38	46.3	724	1	CBBA_BACTJ	Q45730	bacillus th
27	38	46.3	750	1	CBBB_BACTV	Q9ziu5	bacillus th
28	37.5	45.7	208	1	YJ79_AQUAE	O67788	aquifex aeo
29	37.5	45.7	1294	1	RRPO_WCMVM	P09498	white clove
30	37.5	45.7	1294	1	RRPO_WCMVO	P15402	white clove
31	37	45.1	102	1	COXE_CYPCA	O13082	cyprinus ca
32	37	45.1	168	1	Y358_AGR5	Q8uid9	agrobacteri
33	37	45.1	253	1	YK86_XANCP	Q8p8z2	xanthomonas
34	37	45.1	257	1	YJ85_XYLFT	Q87a46	xylella fas
35	37	45.1	257	1	YQ09_XYLEA	Q9paa9	xylella fas
36	37	45.1	291	1	GDBB_WHEAT	P06659	triticum ae
37	37	45.1	337	1	SYW_TREPA	O83640	treponema p
38	37	45.1	385	1	OP23_HAEIN	P46025	haemophilus
39	37	45.1	418	1	CAR5_CANAL	P43094	candida alb
40	37	45.1	446	1	PRRB_MYCLE	O33071	mycobacteri
41	37	45.1	471	1	MM13_BOVIN	O77656	bos taurus
42	37	45.1	474	1	GATA_SCHPO	O13837	schizosacch
43	37	45.1	545	1	PUR9_BIFLO	Q8g6b1	b bifunctio
44	37	45.1	643	1	SLS1_YEAST	P42900	saccharomyc
45	37	45.1	691	1	UVRC_TREPA	O83485	treponema p

# ALIGNMENTS

## RESULT 1

CLH1\_HUMAN

ID CLH1\_HUMAN STANDARD; PRT; 1675 AA.

AC Q00610;

DT 01-DEC-1992 (Rel. 24, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Clathrin heavy chain 1 (CLH-17).

GN CLTC OR CLH17 OR KIAA0034.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Bone marrow;

RX MEDLINE=96051387; PubMed=7584026;

RA Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawarabayasi Y.,

RA Sato S., Nagase T., Seki N., Ishikawa K.-I., Tabata S.;

RT "Prediction of the coding sequences of unidentified human genes. I.

RT The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by

RT analysis of randomly sampled cDNA clones from human immature myeloid

RT cell line KG-1.";

RL DNA Res. 1:27-35(1994).

RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN [3]  
 RP SEQUENCE OF 560-864 FROM N.A.  
 RC TISSUE=Colon;  
 RX MEDLINE=92112210; PubMed=1765375;  
 RA Dodge G.R., Kovalszky I., McBride O.W., Yi H.F., Chu M.L., Saitta B.,  
 RA Stokes D.G., Iozzo R.V.;  
 RT "Human clathrin heavy chain (CLTC): partial molecular cloning,  
 RT expression, and mapping of the gene to human chromosome 17q11-qter."  
 RL Genomics 11:174-178(1991).  
 CC -!- FUNCTION: CLATHRIN IS THE MAJOR PROTEIN OF THE POLYHEDRAL COAT OF  
 CC COATED PITS & VESICLES. TWO DIFFERENT ADAPTOR PROTEIN COMPLEXES  
 CC LINK THE CLATHRIN LATTICE EITHER TO THE PLASMA MEMBRANE OR TO THE  
 CC TRANS GOLGI NETWORK.  
 CC -!- SUBUNIT: Clathrin triskelions, composed of 3 heavy chains and 3  
 CC light chains, are the basic subunits of the clathrin coat. In the  
 CC presence of light chains, hub assembly is influenced by both the  
 CC pH and the concentration of calcium.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic face of coated pits and  
 CC vesicles.  
 CC -!- SIMILARITY: Belongs to the clathrin heavy chain family.  
 CC -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;

WWW="<http://www.infobiogen.fr/services/chromcancer/Genes/CLTCID360.html>".

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; D21260; BAA04801.1; -.

DR EMBL; BC054489; AAH54489.1; -.  
 DR EMBL; X55878; CAA39363.1; -.  
 DR PIR; A40573; A40573.  
 DR HSSP; P11442; 1BPO.  
 DR Genew; HGNC:2092; CLTC.  
 DR MIM; 118955; -.  
 DR GO; GO:0030118; C:clathrin coat; NAS.  
 DR GO; GO:0005198; F:structural molecule activity; NAS.  
 DR GO; GO:0006886; P:intracellular protein transport; NAS.  
 DR InterPro; IPR008938; ARM.  
 DR InterPro; IPR001473; Clathrin\_propel.  
 DR InterPro; IPR000547; Clathrin\_repeat.  
 DR InterPro; IPR008941; TPR-like.  
 DR Pfam; PF00637; Clathrin; 7.  
 DR Pfam; PF01394; Clathrin\_propel; 7.  
 DR SMART; SM00299; CLH; 7.  
 KW Coated pits.  
 FT DOMAIN 1 479 GLOBULAR TERMINAL DOMAIN.  
 FT DOMAIN 480 523 FLEXIBLE LINKER.  
 FT DOMAIN 524 1675 HEAVY CHAIN ARM.  
 FT DOMAIN 524 634 DISTAL SEGMENT.  
 FT DOMAIN 639 1675 PROXIMAL SEGMENT.  
 FT DOMAIN 449 465 BINDING SITE FOR THE UNCOATING ATPASE,  
 FT INVOLVED IN LATTICE DISASSEMBLY  
 FT (POTENTIAL).  
 FT BINDING 1213 1522 LIGHT CHAIN (BY SIMILARITY).  
 FT DOMAIN 1550 1675 TRIMERIZATION (BY SIMILARITY).  
 FT CONFLICT 560 560 Q -> R (IN REF. 3).  
 FT CONFLICT 817 817 G -> V (IN REF. 3).  
 SQ SEQUENCE 1675 AA; 191614 MW; 6C4F2D54950079E2 CRC64;

Query Match 50.6%; Score 41.5; DB 1; Length 1675;  
 Best Local Similarity 64.3%; Pred. No. 74;  
 Matches 9; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 2 KPFPKLVVEVFPFP 15  
 :|||| |:|| ||  
 Db 241 QPFPKKAVDVF-FP 253

## RESULT 2

CLH\_BOVIN

ID CLH\_BOVIN STANDARD; PRT; 1675 AA.  
 AC P49951;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Clathrin heavy chain.  
 GN CLTC.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;

```

RX  MEDLINE=96028100; PubMed=7585943;
RA  Liu S.-H., Wong M.L., Craik C.S., Brodsky F.M.;
RT  "Regulation of clathrin assembly and trimerization defined using
RT  recombinant triskelion hubs.";
RL  Cell 83:257-267(1995).
CC  -!- FUNCTION: Clathrin is the major protein of the polyhedral coat of
CC      coated pits and vesicles. Two different adaptor protein complexes
CC      link the clathrin lattice either to the plasma membrane or to the
CC      trans Golgi network.
CC  -!- SUBUNIT: Clathrin triskelions, composed of 3 heavy chains and 3
CC      light chains, are the basic subunits of the clathrin coat. In the
CC      presence of light chains, hub assembly is influenced by both the
CC      pH and the concentration of calcium.
CC  -!- SUBCELLULAR LOCATION: Cytoplasmic face of coated pits and
CC      vesicles.
CC  -!- DOMAIN: The C-terminal third of the heavy chains forms the hub of
CC      the triskelion. This region contains the trimerization domain and
CC      the light-chain binding domain involved in the assembly of the
CC      clathrin lattice.
CC  -!- SIMILARITY: Belongs to the clathrin heavy chain family.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; U31757; AAC48524.1; -.
DR  PDB; 1B89; 04-JUN-99.
DR  InterPro; IPR008938; ARM.
DR  InterPro; IPR001473; Clathrin_propel.
DR  InterPro; IPR000547; Clathrin_repeat.
DR  InterPro; IPR008941; TPR-like.
DR  Pfam; PF00637; Clathrin; 7.
DR  Pfam; PF01394; Clathrin_propel; 7.
DR  SMART; SM00299; CLH; 7.
KW  Coated pits; 3D-structure.
FT  DOMAIN      1      479      GLOBULAR TERMINAL DOMAIN.
FT  DOMAIN      480     523      FLEXIBLE LINKER.
FT  DOMAIN      524    1675      HEAVY CHAIN ARM.
FT  DOMAIN      524     634      DISTAL SEGMENT.
FT  DOMAIN      639    1675      PROXIMAL SEGMENT.
FT  DOMAIN      449     465      BINDING SITE FOR THE UNCOATING ATPASE,
FT                                     INVOLVED IN LATTICE DISASSEMBLY
FT                                     (POTENTIAL).
FT  BINDING      1213    1522      LIGHT CHAIN.
FT  DOMAIN      1550    1675      TRIMERIZATION.
SQ  SEQUENCE    1675 AA;  191587 MW;  6C4F2D54801579E2 CRC64;

Query Match          50.6%;  Score 41.5;  DB 1;  Length 1675;
Best Local Similarity 64.3%;  Pred. No. 74;
Matches      9;  Conservative      2;  Mismatches      2;  Indels      1;  Gaps      1;

Qy      2 KPFPKCLKVEVFPPF 15
      :|||| |:|| ||

```

## RESULT 3

## CLH\_RAT

ID CLH\_RAT STANDARD; PRT; 1675 AA.  
AC P11442;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Clathrin heavy chain.  
GN CLTC.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88097376; PubMed=3480512;  
RA Kirchhausen T., Harrison S.C., Chow E.P., Mattaliano R.J.,  
RA Ramachandran K.L., Smart J., Brosius J.;  
RT "Clathrin heavy chain: molecular cloning and complete primary  
RT structure.";  
RL Proc. Natl. Acad. Sci. U.S.A. 84:8805-8809(1987).  
RN [2]  
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-493.  
RX MEDLINE=99043510; PubMed=9827808;  
RA Ter Haar E., Musacchio A., Harrison S.C., Kirchhausen T.;  
RT "Atomic structure of clathrin: a beta propeller terminal domain joins  
RT an alpha zigzag linker.";  
RL Cell 95:563-573(1998).  
CC -!- FUNCTION: CLATHRIN IS THE MAJOR PROTEIN OF THE POLYHEDRAL COAT OF  
CC COATED PITS & VESICLES. TWO DIFFERENT ADAPTOR PROTEIN COMPLEXES  
CC LINK THE CLATHRIN LATTICE EITHER TO THE PLASMA MEMBRANE OR TO THE  
CC TRANS GOLGI NETWORK.  
CC -!- SUBUNIT: Clathrin triskelions, composed of 3 heavy chains and 3  
CC light chains, are the basic subunits of the clathrin coat. In the  
CC presence of light chains, hub assembly is influenced by both the  
CC pH and the concentration of calcium.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic face of coated pits and  
CC vesicles.  
CC -!- PTM: The N-terminus is blocked.  
CC -!- SIMILARITY: Belongs to the clathrin heavy chain family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; J03583; AAA40874.1; -.  
DR PIR; A39941; LRRTH.  
DR PDB; 1BPO; 06-APR-99.  
DR PDB; 1C9I; 07-FEB-00.  
DR PDB; 1C9L; 07-FEB-00.

DR	InterPro; IPR008938; ARM.			
DR	InterPro; IPR001473; Clathrin_propel.			
DR	InterPro; IPR000547; Clathrin_repeat.			
DR	InterPro; IPR008941; TPR-like.			
DR	Pfam; PF00637; Clathrin; 7.			
DR	Pfam; PF01394; Clathrin_propel; 7.			
DR	SMART; SM00299; CLH; 7.			
KW	Coated pits; 3D-structure.			
FT	DOMAIN	1	479	LOBULAR TERMINAL DOMAIN.
FT	DOMAIN	480	523	FLEXIBLE LINKER.
FT	DOMAIN	524	1675	HEAVY CHAIN ARM.
FT	DOMAIN	524	634	DISTAL SEGMENT.
FT	DOMAIN	639	1675	PROXIMAL SEGMENT.
FT	DOMAIN	449	465	BINDING SITE FOR THE UNCOATING ATPASE,
FT				INVOLVED IN LATTICE DISASSEMBLY
FT				(POTENTIAL).
FT	BINDING	1213	1522	LIGHT CHAIN (BY SIMILARITY).
FT	DOMAIN	1550	1675	TRIMERIZATION (BY SIMILARITY).
FT	STRAND	7	14	
FT	HELIX	15	18	
FT	TURN	19	19	
FT	TURN	22	23	
FT	TURN	27	29	
FT	STRAND	30	34	
FT	TURN	35	36	
FT	STRAND	37	42	
FT	TURN	45	46	
FT	STRAND	49	54	
FT	TURN	55	56	
FT	TURN	58	59	
FT	STRAND	62	65	
FT	STRAND	70	73	
FT	STRAND	80	84	
FT	TURN	85	86	
FT	STRAND	87	92	
FT	TURN	93	96	
FT	STRAND	97	103	
FT	STRAND	110	115	
FT	TURN	116	117	
FT	STRAND	118	122	
FT	STRAND	126	131	
FT	STRAND	142	143	
FT	HELIX	146	148	
FT	TURN	149	150	
FT	STRAND	152	158	
FT	TURN	160	161	
FT	STRAND	164	173	
FT	TURN	174	175	
FT	STRAND	176	185	
FT	TURN	187	188	
FT	STRAND	191	195	
FT	STRAND	198	204	
FT	TURN	207	208	
FT	STRAND	213	222	
FT	TURN	223	224	
FT	STRAND	225	232	
FT	TURN	238	239	



FT	STRAND	246	249
FT	TURN	254	255
FT	TURN	257	258
FT	STRAND	261	267
FT	TURN	268	271
FT	STRAND	272	277
FT	TURN	278	279
FT	STRAND	281	286
FT	TURN	287	289
FT	STRAND	292	297
FT	STRAND	303	309
FT	TURN	310	313
FT	STRAND	314	319
FT	TURN	320	321
FT	STRAND	323	329
FT	TURN	331	333
FT	HELIX	334	340
FT	TURN	341	341
FT	HELIX	345	355
FT	TURN	356	356
FT	HELIX	361	373
FT	TURN	374	375
FT	HELIX	377	386
FT	HELIX	388	390
FT	TURN	391	392
FT	HELIX	395	402
FT	TURN	403	403
FT	HELIX	413	424
FT	STRAND	427	427
FT	HELIX	430	441
FT	TURN	442	443
FT	HELIX	445	453
FT	TURN	454	455
FT	STRAND	457	457
FT	HELIX	461	468
FT	TURN	469	470
FT	HELIX	472	479
FT	TURN	480	483
FT	HELIX	486	490
FT	TURN	491	492
SQ	SEQUENCE	1675 AA;	191598 MW; C10F54C7ED8C5A61 CRC64;

Query Match 50.6%; Score 41.5; DB 1; Length 1675;  
 Best Local Similarity 64.3%; Pred. No. 74;  
 Matches 9; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

Qy 2 KPFPKLVKVEVFPFP 15  
 :|||| |:|| ||  
 Db 241 QPFPKKAVDVF-FP 253

RESULT 4  
 YFHE\_SCHPO  
 ID YFHE\_SCHPO STANDARD; PRT; 398 AA.  
 AC O42851; P78886;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Hypothetical protein C23A1.14c in chromosome I.  
 GN SPAC23A1.14C.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomyces.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe."  
 RL Nature 415:871-880(2002).

RN [2]  
 RP SEQUENCE OF 141-286 FROM N.A.  
 RC STRAIN=PR745;  
 RX MEDLINE=98162722; PubMed=9501991;  
 RA Yoshioka S., Kato K., Nakai K., Okayama H., Nojima H.;  
 RT "Identification of open reading frames in Schizosaccharomyces pombe  
 RT cDNAs."  
 RL DNA Res. 4:363-369(1997).  
 CC -!- COFACTOR: Pyridoxal phosphate (By similarity).  
 CC -!- SIMILARITY: Belongs to the trans-sulfuration enzymes family.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AL021813; CAA16988.1; -.  
 DR EMBL; D89237; BAA13898.1; -.  
 DR HSSP; P00935; 1CS1.  
 DR GeneDB\_SPombe; SPAC23A1.14c; -.  
 DR InterPro; IPR000277; Cys\_Met\_Meta\_PP.  
 DR Pfam; PF01053; Cys\_Met\_Meta\_PP; 1.  
 DR PROSITE; PS00868; CYS\_MET\_METAB\_PP; FALSE\_NEG.  
 KW Hypothetical protein; Lyase; Pyridoxal phosphate.  
 FT BINDING 212 212 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 FT CONFLICT 268 268 K -> E (IN REF. 2).  
 FT CONFLICT 278 278 Q -> L (IN REF. 2).  
 SQ SEQUENCE 398 AA; 43284 MW; 651C2BCEF59BCEA7 CRC64;

Query Match 50.0%; Score 41; DB 1; Length 398;  
 Best Local Similarity 57.1%; Pred. No. 22;  
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 LKPFPKLKVEVFPF 14  
 | : | | | | | : | : | :  
 Db 52 LQPF TKLAEE DFPY 65

# RESULT 5

## YA60\_METJA

ID YA60\_METJA STANDARD; PRT; 537 AA.  
 AC Q58460;  
 DT 15-MAR-2004 (Rel. 43, Created)  
 DT 15-MAR-2004 (Rel. 43, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Hypothetical protein MJ1060.  
 GN MJ1060.  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanocaldococcaceae; Methanocaldococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 RT jannaschii.";  
 RL Science 273:1058-1073(1996).

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch).

CC -----  
DR EMBL; U67549; AAB99072.1; -.  
DR PIR; C64432; C64432.  
DR TIGR; MJ1060; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 537 AA; 65989 MW; 83E8A0C63B6D0837 CRC64;

Query Match 50.0%; Score 41; DB 1; Length 537;  
Best Local Similarity 75.0%; Pred. No. 29;  
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 FPKLKVEVFPFP 15  
||| || ||| |  
Db 381 FPKDKVIVFPDP 392

#### RESULT 6

##### TLR4\_RAT

ID TLR4\_RAT STANDARD; PRT; 835 AA.  
AC Q9QX05;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Toll-like receptor 4 precursor (Toll4).  
GN TLR4.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Sprague-Dawley; TISSUE=Heart;  
RX MEDLINE=99362487; PubMed=10430608;  
RA Frantz S., Kobzik L., Kim Y.-D., Fukazawa R., Medzhitov R., Lee R.T.,  
RA Kelly R.A.;  
RT "Toll4 (TLR4) expression in cardiac myocytes in normal and failing  
RT myocardium.";  
RL J. Clin. Invest. 104:271-280(1999).  
CC -!- FUNCTION: Cooperates with LY96 and CD14 to mediate the innate  
CC immune response to bacterial lipopolysaccharide (LPS). Acts via  
CC MyD88, TIRAP and TRAF6, leading to NF-kappa-B activation, cytokine  
CC secretion and the inflammatory response (By similarity).  
CC -!- SUBUNIT: Belongs to the lipopolysaccharide (LPS) receptor, a  
CC multi-protein complex containing at least CD14, LY96 and TLR4.  
CC Binds LY96 via the extracellular domain. Binds MyD88 and TIRAP via  
CC their respective TIR domains (By similarity).  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).  
CC -!- SIMILARITY: Belongs to the Toll-like receptor family.  
CC -!- SIMILARITY: Contains 1 TIR domain.  
CC -!- SIMILARITY: Contains 17 leucine-rich (LRR) repeats.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AF057025; AAC13313.1; -.  
 DR HSSP; O60603; 1FYW.  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR000483; LRR\_Cterm.  
 DR InterPro; IPR000157; TIR.  
 DR Pfam; PF00560; LRR; 8.  
 DR Pfam; PF01463; LRRCT; 1.  
 DR Pfam; PF01582; TIR; 1.  
 DR SMART; SM00082; LRRCT; 1.  
 DR SMART; SM00255; TIR; 1.  
 DR PROSITE; PS50104; TIR; 1.  
 KW Receptor; Immune response; Inflammatory response; Signal;  
 KW Transmembrane; Repeat; Leucine-rich repeat; Glycoprotein.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 26 835 TOLL-LIKE RECEPTOR 4.  
 FT DOMAIN 26 638 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 639 659 POTENTIAL.  
 FT DOMAIN 660 835 CYTOPLASMIC (POTENTIAL).  
 FT REPEAT 32 52 LRR 1.  
 FT REPEAT 53 75 LRR 2.  
 FT REPEAT 76 99 LRR 3.  
 FT REPEAT 100 123 LRR 4.  
 FT REPEAT 148 172 LRR 5.  
 FT REPEAT 173 196 LRR 6.  
 FT REPEAT 201 224 LRR 7.  
 FT REPEAT 227 251 LRR 8.  
 FT REPEAT 305 330 LRR 9.  
 FT REPEAT 370 393 LRR 10.  
 FT REPEAT 396 419 LRR 11.  
 FT REPEAT 420 443 LRR 12.  
 FT REPEAT 468 492 LRR 13.  
 FT REPEAT 493 516 LRR 14.  
 FT REPEAT 518 540 LRR 15.  
 FT REPEAT 542 563 LRR 16.  
 FT REPEAT 565 589 LRR 17.  
 FT DOMAIN 670 816 TIR.  
 FT CARBOHYD 34 34 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 43 43 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 75 75 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 204 204 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 237 237 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 248 248 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 281 281 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 307 307 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 492 492 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 524 524 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 572 572 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 575 575 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 622 622 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 835 AA; 96071 MW; DF5E16A30851E3A0 CRC64;

Query Match

49.4%; Score 40.5; DB 1; Length 835;

Best Local Similarity 64.3%; Pred. No. 54;  
Matches 9; Conservative 1; Mismatches 1; Indels 3; Gaps 1;

Qy 1 LKPFPKLKVEVFPF 14  
|||||| : ||  
Db 340 LKPFPKLSL---PF 350

RESULT 7

COBS\_METMA

ID COBS\_METMA STANDARD; PRT; 280 AA.  
AC Q8PVB4;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Cobalamin synthase (EC 2.-.-.-).  
GN COBS OR MM2057.  
OS Methanosarcina mazei (Methanosarcina frisia).  
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;  
OC Methanosarcinaceae; Methanosarcina.  
OX NCBI\_TaxID=2209;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Goe1 / Go1 / ATCC BAA-199 / DSM 3647 / OCM 88;  
RX MEDLINE=22120827; PubMed=12125824;  
RA Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,  
RA Martinez-Arias R., Henne A., Wiezer A., Baeumer S., Jacobi C.,  
RA Brueggemann H., Lienard T., Christmann A., Boemecke M., Steckel S.,  
RA Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,  
RA Fritz H.-J., Gottschalk G.;  
RT "The genome of Methanosarcina mazei: evidence for lateral gene  
RT transfer between Bacteria and Archaea."  
RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).  
CC -!- FUNCTION: Joins Ado-cobinamide-GDP and alpha-ribazole to generate  
CC adenosylcobalamin (Ado-cobalamin) (By similarity).  
CC -!- CATALYTIC ACTIVITY: GDP-cobinamide + alpha-ribazole = cobalamin +  
CC GMP.  
CC -!- PATHWAY: Cobalamin biosynthesis; last step.  
CC -!- SIMILARITY: Belongs to the cobS family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE013445; AAM31753.1; -.  
DR HAMAP; MF\_00719; -; 1.  
DR InterPro; IPR003805; CobS\_synth.  
DR InterPro; IPR001411; TCR\_TetB.  
DR Pfam; PF02654; CobS; 1.  
DR PRINTS; PR01036; TCRTETB.  
DR TIGRFAMs; TIGR00317; cobS; 1.  
KW Cobalamin biosynthesis; Transferase; Complete proteome.  
SQ SEQUENCE 280 AA; 29618 MW; D8C06C3BCF5CA798 CRC64;

Query Match 48.8%; Score 40; DB 1; Length 280;  
 Best Local Similarity 50.0%; Pred. No. 22;  
 Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 2 KPFPKLVKEVFP 13  
 |||:|:|:|  
 Db 170 KPLPRLKEQTYP 181

# RESULT 8

## TYSY\_YEAST

ID TYSY\_YEAST STANDARD; PRT; 304 AA.  
 AC P06785; Q12694;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Thymidylate synthase (EC 2.1.1.45) (TS) (TSase).  
 GN TMP1 OR CDC21 OR YOR074C OR YOR29-25.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87165970; PubMed=3031048;  
 RA Taylor G.R., Lagosky P.A., Storms R.K., Haynes R.H.;  
 RT "Molecular characterization of the cell cycle-regulated thymidylate  
 RT synthase gene of Saccharomyces cerevisiae.";  
 RL J. Biol. Chem. 262:5298-5307(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97279235; PubMed=9133743;  
 RA Valens M., Bohn C., Daignan-Fornier B., Dang V., Bolotin-Fukuhara M.;  
 RT "The sequence of a 54.7 kb fragment of yeast chromosome XV reveals  
 RT the presence of two tRNAs and 24 new open reading frames.";  
 RL Yeast 13:379-390(1997).  
 RN [3]  
 RP SEQUENCE OF 1-38 FROM N.A.  
 RX MEDLINE=89096830; PubMed=3062362;  
 RA McIntosh E.M., Ord R.W., Storms R.K.;  
 RT "Transcriptional regulation of the cell cycle-dependent thymidylate  
 RT synthase gene of Saccharomyces cerevisiae.";  
 RL Mol. Cell. Biol. 8:4616-4624(1988).  
 CC -!- FUNCTION: REQUIRED FOR BOTH NUCLEAR AND MITOCHONDRIAL DNA  
 CC SYNTHESIS.  
 CC -!- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + dUMP =  
 CC dihydrofolate + dTMP.  
 CC -!- PATHWAY: Deoxyribonucleotide biosynthesis.  
 CC -!- SUBUNIT: Homodimer.  
 CC -!- SIMILARITY: Belongs to the thymidylate synthase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; J02706; AAA60940.1; -.  
DR EMBL; Z74982; CAA99267.1; ALT\_SEQ.  
DR EMBL; Z70678; CAA94559.1; -.  
DR EMBL; M29100; AAA35159.1; -.  
DR PIR; S66957; YXBYT.  
DR HSSP; P04818; 1HW4.  
DR GermOnline; 143662; -.  
DR SGD; S0005600; CDC21.  
DR GO; GO:0005634; C:nucleus; IDA.  
DR GO; GO:0004799; F:thymidylate synthase activity; IDA.  
DR InterPro; IPR000398; Thymidylat\_synth.  
DR Pfam; PF00303; thymidylat\_synt; 1.  
DR PRINTS; PR00108; THYMDSNTHASE.  
DR ProDom; PD001180; Thymidylat\_synt; 1.  
DR PROSITE; PS00091; THYMIDYLATE\_SYNTHASE; 1.  
KW Transferase; Methyltransferase; Nucleotide biosynthesis.  
FT ACT\_SITE 177 177 BY SIMILARITY.  
SQ SEQUENCE 304 AA; 35047 MW; 0C514BEDB8574510 CRC64;

Query Match 48.8%; Score 40; DB 1; Length 304;  
Best Local Similarity 66.7%; Pred. No. 24;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KPFPKLKVE 10  
:|||||::  
Db 265 RPFPLKLIK 273

#### RESULT 9

##### PGMP\_PEA

ID PGMP\_PEA STANDARD; PRT; 626 AA.  
AC Q9SM59;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Phosphoglucumutase, chloroplast precursor (EC 5.4.2.2) (Glucose  
DE phosphomutase) (PGM).  
GN PGMP OR RUG3.  
OS Pisum sativum (Garden pea).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.  
OX NCBI\_TaxID=3888;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BC1; TISSUE=Cotyledon;  
RX MEDLINE=20223733; PubMed=10759514;  
RA Harrison C.J., Mould R.M., Leech M.J., Johnson S.A., Turner L.,  
RA Schreck S.L., Baird K.M., Jack P.L., Rawsthorne S., Hedley C.L.,  
RA Wang T.L.;  
RT "The rug3 locus of pea encodes plastidial phosphoglucumutase."  
RL Plant Physiol. 122:1187-1192(2000).  
CC -!- FUNCTION: This enzyme participates in both the breakdown and  
CC synthesis of glucose (By similarity).



CC -!- CATALYTIC ACTIVITY: Alpha-D-glucose 1-phosphate = alpha-D-glucose  
 CC 6-phosphate.  
 CC -!- COFACTOR: Magnesium (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Chloroplast.  
 CC -!- SIMILARITY: Belongs to the phosphohexose mutase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AJ250770; CAB60128.1; -.  
 DR HSSP; P00949; 3PMG.  
 DR InterPro; IPR005841; PG\_PMM\_mutase.  
 DR InterPro; IPR005844; PG\_PMM\_ABAI.  
 DR InterPro; IPR005845; PG\_PMM\_ABAII.  
 DR InterPro; IPR005846; PG\_PMM\_ABAIII.  
 DR InterPro; IPR005843; PG\_PMM\_C.  
 DR Pfam; PF00408; PGM\_PMM; 1.  
 DR Pfam; PF02878; PGM\_PMM\_I; 1.  
 DR Pfam; PF02879; PGM\_PMM\_II; 1.  
 DR Pfam; PF02880; PGM\_PMM\_III; 1.  
 DR PRINTS; PR00509; PGM\_PMM.  
 DR PROSITE; PS00710; PGM\_PMM; 1.  
 KW Isomerase; Phosphorylation; Magnesium; Chloroplast; Transit peptide.  
 FT TRANSIT 1 66 CHLOROPLAST (POTENTIAL).  
 FT CHAIN 67 626 PHOSPHOGLUCOMUTASE.  
 FT ACT\_SITE 184 184 PHOSPHOSERINE INTERMEDIATE  
 FT (BY SIMILARITY).  
 SQ SEQUENCE 626 AA; 68574 MW; B820E069AFA0D34E CRC64;

Query Match 48.8%; Score 40; DB 1; Length 626;  
 Best Local Similarity 63.6%; Pred. No. 49;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 FPKLKVEVFPPF 14  
 || ||: |||  
 Db 33 FPSFKVQNFPF 43

#### RESULT 10

##### TYSY\_MOUSE

ID TYSY\_MOUSE STANDARD; PRT; 307 AA.  
 AC P07607;  
 DT 01-APR-1988 (Rel. 07, Created)  
 DT 01-APR-1988 (Rel. 07, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Thymidylate synthase (EC 2.1.1.45) (TS) (TSase).  
 GN TYMS.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=88174353; PubMed=3444407;  
 RA Perryman S.M., Rossana C., Deng T., Vanin E.F., Johnson L.F.;  
 RT "Sequence of a cDNA for mouse thymidylate synthase reveals striking  
 RT similarity with the prokaryotic enzyme.";  
 RL Mol. Biol. Evol. 3:313-321(1986).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87057259; PubMed=3782103;  
 RA Deng T., Li D., Jenh C.-H., Johnson L.F.;  
 RT "Structure of the gene for mouse thymidylate synthase. Locations of  
 RT introns and multiple transcriptional start sites.";  
 RL J. Biol. Chem. 261:16000-16005(1986).  
 RN [3]  
 RP SEQUENCE OF 236-265 FROM N.A.  
 RX MEDLINE=89128436; PubMed=2915925;  
 RA Deng T., Li Y., Johnson L.F.;  
 RT "Thymidylate synthase gene expression is stimulated by some (but not  
 RT all) introns.";  
 RL Nucleic Acids Res. 17:645-658(1989).  
 CC -!- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + dUMP =  
 CC dihydrofolate + dTMP.  
 CC -!- PATHWAY: Deoxyribonucleotide biosynthesis.  
 CC -!- SUBUNIT: Homodimer.  
 CC -!- SIMILARITY: Belongs to the thymidylate synthase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M13019; AAA40439.1; -.  
 DR EMBL; M13352; AAA40444.1; -.  
 DR EMBL; J02617; AAA40444.1; JOINED.  
 DR EMBL; M13347; AAA40444.1; JOINED.  
 DR EMBL; M13348; AAA40444.1; JOINED.  
 DR EMBL; M13349; AAA40444.1; JOINED.  
 DR EMBL; M13350; AAA40444.1; JOINED.  
 DR EMBL; M13351; AAA40444.1; JOINED.  
 DR EMBL; X14489; CAA32651.1; -.  
 DR PIR; A26323; YXMST.  
 DR HSSP; P45352; 1RTS.  
 DR MGD; MGI:98878; Tyms.  
 DR InterPro; IPR000398; Thymidylat\_synth.  
 DR Pfam; PF00303; thymidylat\_synt; 1.  
 DR PRINTS; PR00108; THYMDSNTHASE.  
 DR ProDom; PD001180; Thymidylat\_synt; 1.  
 DR PROSITE; PS00091; THYMIDYLATE\_SYNTHASE; 1.  
 KW Transferase; Methyltransferase; Nucleotide biosynthesis.  
 FT ACT\_SITE 189 189 BY SIMILARITY.  
 SQ SEQUENCE 307 AA; 34958 MW; E4930618C487FD5E CRC64;

Query Match 47.6%; Score 39; DB 1; Length 307;  
 Best Local Similarity 75.0%; Pred. No. 36;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KPFPKLKV 9  
 :|||||:  
 Db 268 RPFPLKI 275

RESULT 11

GDB2\_WHEAT

ID GDB2\_WHEAT STANDARD; PRT; 327 AA.  
 AC P08453;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 01-NOV-1990 (Rel. 16, Last annotation update)  
 DE Gamma-gliadin precursor.  
 OS Triticum aestivum (Wheat).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;  
 OC Triticeae; Triticum.  
 OX NCBI\_TaxID=4565;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Sugiyama T., Rafalski A., Soell D.;  
 RT "The nucleotide sequence of a wheat gamma-gliadin genomic clone."  
 RL Plant Sci. 44:205-209(1986).  
 CC -!- FUNCTION: Gliadin is the major seed storage protein in wheat.  
 CC -!- MISCELLANEOUS: THE GAMMA-GLIADINS CAN BE DIVIDED INTO 3 HOMOLOGY  
 CC CLASSES. SEQUENCE DIVERGENCE BETWEEN THE CLASSES IS DUE TO  
 CC SINGLE-BASE SUBSTITUTIONS & TO DUPLICATIONS OR DELETIONS WITHIN OR  
 CC NEAR DIRECT REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M16064; AAA34289.1; -.  
 DR PIR; JS0402; JS0402.  
 DR InterPro; IPR003612; AAI.  
 DR InterPro; IPR001954; Glia\_glutenin.  
 DR Pfam; PF00234; tryp\_alpha\_amyl; 1.  
 DR PRINTS; PR00208; GLIADGLUTEN.  
 DR SMART; SM00499; AAI; 1.  
 KW Seed storage protein; Repeat; Signal; Multigene family.  
 FT SIGNAL 1 19  
 FT CHAIN 20 327 GAMMA-GLIADIN.  
 SQ SEQUENCE 327 AA; 37122 MW; E27FEB9DA8BDFCCB CRC64;

Query Match 47.6%; Score 39; DB 1; Length 327;  
 Best Local Similarity 50.0%; Pred. No. 38;  
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 2 KPFPKLKVEVFPFP 15  
 :|||:|: |||

## RESULT 12

## CD22\_PONPY

ID CD22\_PONPY STANDARD; PRT; 330 AA.  
AC Q9N1E3;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE B-cell receptor CD22 precursor (Siglec-2) (Fragment).  
GN CD22.  
OS Pongo pygmaeus (Orangutan).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.  
OX NCBI\_TaxID=9600;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20187579; PubMed=10722703;  
RA Brinkman-Van der Linden E.C.M., Sjoberg E.R., Juneja L.R.,  
RA Crocker P.R., Varki N., Varki A.;  
RT "Loss of N-glycolylneuraminic acid in human evolution: implications  
RT for sialic acid recognition by siglecs.";  
RL J. Biol. Chem. 275:8633-8640(2000).  
CC -!- FUNCTION: Mediates B-cell B-cell interactions. May be involved in  
CC the localization of B-cells in lymphoid tissues. Binds sialylated  
CC glycoproteins; one of which is CD45. Preferentially binds to  
CC alpha2,6-linked sialic acid (By similarity). Upon ligand induced  
CC tyrosine phosphorylation in the immune response seems to be  
CC involved in regulation of B cell antigen receptor signaling. Plays  
CC a role in positive regulation through interaction with Src family  
CC tyrosine kinases and may also act as an inhibitory receptor by  
CC recruiting cytoplasmic phosphatases via their SH2 domains that  
CC block signal transduction through dephosphorylation of signaling  
CC cascade key molecules.  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- SIMILARITY: Belongs to the immunoglobulin superfamily. SIGLEC  
CC (sialic acid binding Ig-like lectin) family.  
CC -!- SIMILARITY: Contains at least 2 immunoglobulin-like C2-type  
CC domains.  
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF199418; AAF44617.1; -.  
DR InterPro; IPR007110; Ig-like.  
DR InterPro; IPR003598; Ig\_c2.  
DR InterPro; IPR003006; Ig\_MHC.  
DR Pfam; PF00047; ig; 2.  
DR SMART; SM00408; IGc2; 1.  
DR PROSITE; PS50835; IG\_LIKE; 2.

DR PROSITE; PS00290; IG\_MHC; FALSE\_NEG.  
KW Cell adhesion; Lectin; Signal; Glycoprotein; Immunoglobulin domain;  
KW Repeat.  
FT SIGNAL 1 17 POTENTIAL.  
FT CHAIN 18 330 B-CELL RECEPTOR CD22.  
FT DOMAIN 18 >330 EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 18 136 IG-LIKE V-TYPE.  
FT DOMAIN 141 233 IG-LIKE C2-TYPE 1.  
FT DOMAIN 240 324 IG-LIKE C2-TYPE 2.  
FT DISULFID 37 165 BY SIMILARITY.  
FT DISULFID 42 100 BY SIMILARITY.  
FT DISULFID 159 217 BY SIMILARITY.  
FT DISULFID 263 307 BY SIMILARITY.  
FT NON\_TER 330 330  
SQ SEQUENCE 330 AA; 37257 MW; E7F67002FD5F5381 CRC64;

Query Match 47.6%; Score 39; DB 1; Length 330;  
Best Local Similarity 61.5%; Pred. No. 38;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 LKPFPKLKVEVEFP 13  
:| |||:| |  
Db 236 VKHTPKLKIEVNP 248

# RESULT 13

## YEBU\_ECOLI

ID YEBU\_ECOLI STANDARD; PRT; 479 AA.  
AC P76273; O07980;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein yebU.  
GN YEBU OR B1835.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12."  
RL Science 277:1453-1474(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97251358; PubMed=9097040;  
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,  
RA Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,  
RA Mizobuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y.,  
RA Nashimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y.,  
RA Sivasundaram S., Tagami H., Takeda J., Takemoto K., Wada C.,

RA Yamamoto Y., Horiuchi T.;  
 RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome  
 RT corresponding to the 40.1-50.0 min region on the linkage map.";  
 RL DNA Res. 3:379-392(1996).  
 CC -!- SIMILARITY: BELONGS TO THE SUN (BACTERIAL) / NUCLEOLAR PROTEIN  
 CC NOL1/NOP2 (EUKARYOTES) FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE000278; AAC74905.1; ALT\_INIT.  
 DR EMBL; D90827; BAA15648.1; ALT\_INIT.  
 DR EcoGene; EG14023; yebU.  
 DR InterPro; IPR000051; SAM\_bind.  
 DR InterPro; IPR001678; Sun\_Nop1/Nop2.  
 DR Pfam; PF01189; Noll\_Nop2\_Sun; 1.  
 DR TIGRFAMs; TIGR00446; nop2p; 1.  
 DR PROSITE; PS01153; NOL1\_NOP2\_SUN; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 479 AA; 53227 MW; 30D2B6AD01FCFF3E CRC64;

Query Match 47.6%; Score 39; DB 1; Length 479;  
 Best Local Similarity 66.7%; Pred. No. 55;  
 Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 PFPKLVVEVFPF 14  
 | || || ||  
 Db 318 PAPKYKVGNFPPF 329

#### RESULT 14

##### SSAV\_SALTY

ID SSAV\_SALTY STANDARD; PRT; 681 AA.  
 AC P74856;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Secretion system apparatus protein ssaV.  
 GN SSAV OR STM1414.  
 OS Salmonella typhimurium.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LT2;  
 RX MEDLINE=97285756; PubMed=9140973;  
 RA Hensel M., Shea J.E., Raupach B., Monack D., Falkow S., Gleeson C.,  
 RA Kubo T., Holden D.W.;  
 RT "Functional analysis of ssaJ and the ssaK/U operon, 13 genes encoding  
 RT components of the type III secretion apparatus of Salmonella  
 RT pathogenicity island 2.";

RL Mol. Microbiol. 24:155-167(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
 RA Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
 RT LT2.";  
 RL Nature 413:852-856(2001).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
 CC (Potential).  
 CC -!- SIMILARITY: BELONGS TO THE FHIPEP (FLAGELLA/HR/INVASION PROTEINS  
 CC EXPORT PORE) FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Y09357; CAA70536.1; -.  
 DR EMBL; AE008761; AAL20338.1; -.  
 DR StyGene; SG10719; ssaV.  
 DR InterPro; IPR001712; Bact\_FHIPEP.  
 DR InterPro; IPR006302; HrcV.  
 DR Pfam; PF00771; FHIPEP; 1.  
 DR PRINTS; PR00949; TYPE3IMAPROT.  
 DR TIGRFAMs; TIGR01399; hrcV; 1.  
 DR PROSITE; PS00994; FHIPEP; 1.  
 KW Transport; Protein transport; Inner membrane; Transmembrane;  
 KW Complete proteome.  
 FT TRANSMEM 25 45 POTENTIAL.  
 FT TRANSMEM 48 68 POTENTIAL.  
 FT TRANSMEM 73 93 POTENTIAL.  
 FT TRANSMEM 118 138 POTENTIAL.  
 FT TRANSMEM 206 226 POTENTIAL.  
 FT TRANSMEM 244 264 POTENTIAL.  
 FT TRANSMEM 295 315 POTENTIAL.  
 FT TRANSMEM 409 429 POTENTIAL.  
 SQ SEQUENCE 681 AA; 75321 MW; C9226C9F9A16114A CRC64;

Query Match 47.6%; Score 39; DB 1; Length 681;  
 Best Local Similarity 46.2%; Pred. No. 78;  
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 3 PFPKLVKEVFPFP 15  
 | |:::|| | |  
 Db 384 PLPEVNIEVLPEP 396

RESULT 15

Y044\_UREPA  
 ID Y044\_UREPA STANDARD; PRT; 782 AA.  
 AC Q9PRA1;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein UU044.  
 GN UU044.  
 OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).  
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.  
 OX NCBI\_TaxID=134821;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Serovar 3;  
 RX MEDLINE=20500219; PubMed=11048724;  
 RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,  
 RA Cassell G.H.;  
 RT "The complete sequence of the mucosal pathogen Ureaplasma  
 RT urealyticum.";  
 RL Nature 407:757-762(2000).  
 CC -!- SIMILARITY: STRONG, TO U.PARVUM UU046.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE002104; AAF30449.1; -.  
 KW Hypothetical protein; Transmembrane; Complete proteome.  
 FT TRANSMEM 10 30 POTENTIAL.  
 SQ SEQUENCE 782 AA; 88546 MW; 9200C82C0FAFF11D CRC64;

Query Match 47.6%; Score 39; DB 1; Length 782;  
 Best Local Similarity 57.1%; Pred. No. 90;  
 Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 2 KPFPKLVKVEVFPFP 15  
 || || | : |||  
 Db 78 KPQKPKPQPTFPFP 91

Search completed: August 24, 2004, 15:43:41  
 Job time : 14.0597 secs